



**UNIVERSIDADE NOVA DE LISBOA**

**EXPRESSÃO EM *ESCHERICHIA COLI* DE ANTIGÉNIOS DO  
*CELL FUSING AGENT VIRUS (FLAVIVIRIDAE: FLAVIVIRUS)*  
COMO PROTEÍNA DE FUSÃO**

**SANDRA EUGÉNIA LEITE MARQUES**

**DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE MESTRE EM  
MICROBIOLOGIA MÉDICA**

**NOVEMBRO**

**2012**



**UNIVERSIDADE NOVA DE LISBOA**

**EXPRESSÃO EM *ESCHERICHIA COLI* DE ANTIGÉNIOS DO  
*CELL FUSING AGENT VIRUS (FLAVIVIRIDAE: FLAVIVIRUS)*  
COMO PROTEÍNA DE FUSÃO**

**SANDRA EUGÉNIA LEITE MARQUES**

*Dissertação para a obtenção do grau de Mestre em Microbiologia Médica*

**ORIENTADOR:**

**PROFESSOR DOUTOR RICARDO PARREIRA**

**Unidade de Microbiologia Médica/Grupo de Virologia  
Instituto de Higiene e Medicina Tropical**

**NOVEMBRO**

**2012**

## AGRADECIMENTOS

Chego ao fim do trabalho de estágio e com ele o término do curso, é com sincera satisfação e alegria que agradeço a todos os que, directa ou indirectamente, contribuíram para que se tornasse possível a realização deste trabalho. Aqui deixo o meu mais profundo reconhecimento:

Ao Professor Doutor Ricardo Parreira, orientador desta dissertação, por toda a ajuda que me prestou tornando possível a realização deste trabalho. Não poderia deixar de reconhecer o seu empenho, paciência, apoio, o sentido crítico, e pela disponibilidade demonstrada em todas as fases que levaram à concretização deste trabalho. O meu sincero e sentido: obrigada.

À Professora Doutora Aida Esteves, pela ajuda, pelo incentivo especialmente nos meus momentos de desânimo, pela sua permanente disponibilidade e paciência para resolver algumas etapas desta tese.

Ao Professor Doutor João Piedade, pelas sugestões, pelo apoio e motivação.

À Teresa Veneno, por todo o apoio e disponibilidade que sempre me dedicou.

À Sandra Costa, pelos seus ensinamentos e conselhos preciosos. Não poderia deixar de expressar um agradecimento especial pela sua amizade.

À técnica da Unidade de Virologia, IHMT, D<sup>a</sup> Fernanda, pelo empenhamento e prontidão com que acolheu todos os pedidos “urgentes”.

À Carina e Tatiana, colegas de mestrado e acima de tudo, grandes amigas, pela ajuda, pelo respeito e companheirismo, nas horas mais difíceis que passamos, foram imprescindíveis para que eu chegasse até ao final deste trabalho.

Ao André, pela amizade, companheirismo, partilha de conhecimentos e momentos de convívio.

Por último, aos meus familiares, especialmente, a minha tia Cristina por todo apoio, e que não esquecerei.

*Dedico esta tese ao meu querido pai, com muitas saudades.*

*E à minha mãe.*

## RESUMO

O *Cell Fusing Agent Vírus* (CFAV), considerado como o primeiro “flavivírus específicos de insectos” (ISF), parece estar exclusivamente adaptado aos seus hospedeiros, não replicando em células de vertebrados. Apesar de ter sido identificado há mais de três décadas (1975), a verdade é que muito pouco se conhece sobre a sua biologia. Dado o seu parentesco filogenético com alguns outros flavivírus encontrados naturalmente em mosquitos de diferentes géneros colhidos em diferentes regiões do globo, este vírus poderá ser usado como modelo para o estudo de ISF. No entanto, necessitam do desenvolvimento de ferramentas básicas, tais como clones moleculares ou baterias de soros contendo anticorpos que reconheçam uma ou mais proteínas codificadas pelo genoma viral, produzidas, por exemplo, a partir de antígenos virais produzidos de forma recombinante.

Com este trabalho pretendeu-se a optimização de protocolos que permitiram a expressão e purificação parcial de quatro proteínas [duas proteínas estruturais (C e E) e duas não estruturais (NS3hel e NS5B)] do CFAV em *E. coli*, todas elas produzidas como proteínas de fusão com “caudas” (*tags*) de hexahistidina nos seus extremos carboxilo.

Para a expansão do CFAV foram utilizadas células *Aedes albopictus* (C6/36). Após a realização da extracção do RNA viral e a obtenção de cDNA, procedeu-se amplificação, por RT-PCR, das regiões codificantes das proteínas C, E, NS3hel e NS5B, utilizando *primers* específicos. Os quatro fragmentos de DNA foram independentemente inseridos no vector pJTE1.2/*blunt* usando *E. coli* NovaBlue como hospedeira de clonagem e, posteriormente, inseridos em vectores de expressão pET-28b e pET-29b usando *E. coli* BL21(DE3)pLysS e Rosetta(DE3)pLysS como hospedeiras de expressão.

Após da indução, expressão e purificação das proteínas recombinantes C, E, NS3hel e NS5B, foi confirmada a autenticidade destas proteínas produzidas através do método *Western Blot* com um anticorpo anti-histidina.

## ABSTRACT

The *Cell Fusing Agent virus* (CFAV) considered as the first "insect- specific flavivirus" (ISF) and seems to be uniquely adapted to their hosts, not replicating in vertebrate cells. Although it has been known for more than three decades (1975), the truth is very little is known about its biology. Given its close phylogenetic relationship with other flavivirus naturally circulating in various genera of mosquitoes collected from different regions of the globe, this virus could be used as a model for the study of ISF. However, such studies require the development of experimental basic tools, such as molecular clones or serum batteries containing antibodies that recognize one or more proteins encoded by the viral genome, produced, for example, from viral antigens recombinant produced.

In this work, we carried out the optimization of protocols that allowed the expression and partial purification of four proteins [two structural proteins (C and E) and two nonstructural proteins (NS3hel and NS5B)] CFAV in *E. coli* as fusion protein for c-terminal hexahistidine tags.

For the expansion of the CFAV we used *Aedes albopictus* (C6/36) cells. After completion of the viral RNA extraction and cDNA obtained, amplification of the coding regions of the C, E, NS5B and NS3hel proteins was carried out by RT-PCR using specific primers. The four DNA fragments were independently inserted into the vector pJTE1.2/*blunt* using *E. coli* NovaBlue as cloning host and then inserted into expression vectors pET-28b and pET-29b using *E. coli* BL21(DE3)pLysS and Rosetta(DE3)pLysS as expression host.

After induction, expression and purification of recombinant C, E, NS3hel and NS5B proteins *Western Blot* analyses with an anti-histidine antibody confirmed the authenticity of these proteins produced.

## ÍNDICE GERAL

AGRADECIMENTOS.....	I
RESUMO.....	III
ABSTRACT.....	IV
ÍNDICE GERAL.....	V
ÍNDICE DE FIGURAS.....	IX
ÍNDICE DE TABELAS.....	XII
ABREVIATURAS.....	XIII

### 1. INTRODUÇÃO

1.1. Considerações gerais do género.....	2
1.2. <i>Cell Fusing Agent Virus</i> (CFAV): da sua descoberta à actualidade.....	4
1.3. Ciclo replicativo.....	6
1.4. Proteínas estruturais e não estruturais dos flavivírus.....	7
1.5. Sistemas de expressão de proteínas recombinantes.....	11
1.5.1. <i>Escherichia coli</i> como hospedeira de expressão de proteínas recombinantes.....	14
1.6. O sistema pET.....	15
1.7. Notas finais e objectivos.....	17

### 2. MATERIAL E MÉTODOS

2.1. Expansão do CFAV.....	19
----------------------------	----

<b>2.2.</b>	Extracção de DNA total a partir dos sedimentos celulares.....	19
2.2.1.	Amplificação de fragmentos de DNA por reacção em cadeia da polimerase (PCR).....	20
2.2.2.	Visualização dos produtos de PCR .....	22
<b>2.3.</b>	Extracção do RNA viral.....	22
2.3.1.	Síntese, por transcrição reversa (RT), de DNA complementar (cDNA) a partir de RNA viral.....	23
<b>2.4.</b>	Delineamento de oligonucleótidos sintéticos ( <i>primers</i> ) iniciadores de PCR.....	24
<b>2.5.</b>	Amplificação de sequências virais a partir de cDNA.....	26
<b>2.6.</b>	Clonagem molecular de fragmentos do genoma do CFAV no vector pJET1.2/ <i>blunt</i> .....	26
2.6.1.	Preparação de células competentes de <i>E. coli</i> e sua transformação por choque térmico.....	27
2.6.2.	Extracção do DNA plasmídico por lise alcalina.....	28
2.6.3.	Digestão enzimática de DNA plasmídico.....	29
2.6.4.	Análise de moléculas recombinantes de DNA plasmídico através da amplificação, por PCR, dos fragmentos clonados.....	30
2.6.5	Purificação dos produtos de PCR.....	31
<b>2.7.</b>	Preparação dos vectores de expressão pET-28b e pET-29b.....	31
2.7.1.	Digestão enzimática dos produtos de PCR purificados e dos vectores de expressão pET-28b e pET-29b.....	32
<b>2.8.</b>	Clonagem molecular de fragmentos do genoma do CFAV nos vectores de expressão pET-28b e pET-29b.....	33
2.8.1.	Purificação de DNA plasmídico.....	34



2.8.2. Sequenciação de DNA.....	34
<b>2.9. Expressão de proteínas recombinantes.....</b>	<b>35</b>
2.9.1. Indução da expressão de proteínas recombinantes e avaliação da sua toxicidade através da monitorização do crescimento bacteriano em meio líquido.....	36
2.9.2. Efeito da temperatura e concentração de indutor na expressão proteica.....	37
2.9.3. Preparação de extractos de proteínas solúveis totais e sua análise por SDS-PAGE.....	37
<b>2.10. Purificação de proteínas recombinantes com “caudas” de hexahistidina.....</b>	<b>38</b>
2.10.1 Purificação de proteínas em condições desnaturantes.....	39
2.10.2. Purificação de proteínas em condições nativas.....	39
<b>2.11. Detecção de proteínas por <i>Western Blot</i>.....</b>	<b>40</b>
2.11.1. Transferência de proteínas para a membrana de nitrocelulose..	40
2.11.2. Bloqueio e incubação com anticorpo.....	41
<b>2.12 Quantificação das proteínas.....</b>	<b>41</b>
 <b>3. RESULTADOS</b>	
 <b>3.1. Expansão do CFAV em cultura de células C6/36.....</b>	<b>43</b>
<b>3.2. Amplificação de fragmentos de DNA a partir dos sedimentos de células C6/36, por reacção em cadeia da polimerase (PCR) .....</b>	<b>44</b>
<b>3.3. Amplificação dos fragmentos do genoma do CFAV a partir de DNA complementar (cDNA).....</b>	<b>46</b>

3.3.1.	Clonagem molecular de fragmentos do genoma do CFAV no vector pJET1.2/ <i>blunt</i> .....	47
3.4.	Clonagem das regiões codificantes das proteínas C, E, NS3hel e NS5B em vectores de expressão proteica (pET-28b e pET-29b).....	49
3.5.	Expressão de proteínas recombinantes.....	51
3.5.1.	Efeito da indução da expressão de proteínas heterólogas sobre o crescimento de cultura de <i>E. coli</i> BL21(DE3)pLysS e Rosetta(DE3)pLysS.....	51
3.5.2.	Efeito da temperatura e concentração de indutor na expressão de proteínas.....	58
3.5.3.	Verificação da solubilidade proteínas heterólogas.....	61
3.5.4.	Purificação de proteínas recombinantes com “caudas” de hexahistidina.....	62
3.5.5.	Detecção de proteínas por <i>Western Blot</i> .....	64
4	<b>DISCUSSÃO DE RESULTADOS</b>	67
5	<b>CONCLUSÕES E PERSPECTIVAS FUTURAS</b>	79
6	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	82

# ÍNDICE DE FIGURAS

## 1. INTRODUÇÃO

<b>Figura 1.1.</b>	Representação esquemática da organização do genoma e processamento da poliproteína dos flavivírus.....	5
<b>Figura 1.2.</b>	Ciclo replicativo dos flavivírus.....	7
<b>Figura 1.3.</b>	Estratégia para a regular a expressão de genes clonados dentro de um vector pET.....	16

## 2. MATERIAL E MÉTODOS

<b>Figura 2.1.</b>	Esquema representativo do vector pJET1.2/ <i>blunt</i> .....	27
<b>Figura 2.2.</b>	Esquema representativo dos mapas dos vectores pET-28a e pET -29a.....	32

## 3. RESULTADOS

<b>Figura 3.1.</b>	Observação ao microscópico óptico de células C6/36 não infectadas (A; x100) ou infectadas com o CFAV (B; x400).....	43
<b>Figura 3.2.</b>	Separação por electroforese em gel de agarose dos produtos de PCR obtidos usando <i>primers</i> complementares à região ITS1 do operão de rRNA e DNA extraído de células C6/36 infectadas, ou não, com o CFAV.....	44
<b>Figura 3.3.</b>	Observação dos produtos de PCR resultantes da amplificação dos fragmentos de DNA correspondentes as regiões C, E, NS3hel e NS5B (extraído dos sedimentos celulares de C6/36 infectadas com o CFAV), utilizando os <i>primers</i> específicos e	

	após a sua separação por electroforese em gel de agarose.....	45
<b>Figura 3.4.</b>	Separação por electroforese em gel de agarose dos fragmentos de DNA amplificados por PCR a partir de cDNA produzido por transcrição reversa (RT) a partir de RNA total extraído de sobrenadante de cultura de células C6/36 infectadas pelo CFAV.....	46
<b>Figura 3.5.</b>	Análise do padrão de migração, em gel de agarose, do DNA plasmídico de alguns dos clones bacterianos obtidos na sequências de clonagem de fragmentos de DNA referentes às regiões C, E, NS3hel e NS5B.....	48
<b>Figura 3.6.</b>	Análise do perfil de digestão resultante da hidrólise, pelos enzimas <i>VspI</i> e <i>NcoI</i> de potenciais plasmídeos recombinantes contendo o inserto correspondente a região codificante da proteína NS3hel.....	50
<b>Figura 3.7.</b>	Curvas de crescimento em meio LB, a 30 e 37°C de culturas de <i>E.coli</i> Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-28b::C e pET-28b::E, e induzidas a uma densidade óptica (D.O. <sub>600 nm</sub> ) de 0.5 com concentração de 1 mM de IPTG.....	53
<b>Figura 3.8.</b>	Curvas de crescimento em meio LB, a 30 e 37°C de culturas de <i>E. coli</i> Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-29b::NS3hel e pET-29b::NS5B, e induzidas a uma densidade óptica (D.O. <sub>600 nm</sub> ) de 0.5 com concentração de 1 mM de IPTG.....	54
<b>Figura 3.9.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais, 4 horas após indução com 1 mM de IPTG a 30 e 37°C.....	55
<b>Figura 3.10.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais, 4 horas	

	após indução com 1 mM de IPTG a 30 e 37°C.....	57
<b>Figura 3.11.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais, 4 horas após indução com 1 mM de IPTG a 30 e 37 °C.....	57
<b>Figura 3.12.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) dos extractos protéicos totais induzidos.....	59
<b>Figura 3.13.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) dos extractos proteicos totais induzidos com 0,4 e 1 mM a 37°C durante 4 horas.....	60
<b>Figura 3.14.</b>	Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de alíquotas das fracções solúveis (Fs) e insolúveis (Fi) obtidas a partir de extractos onde foi induzida a produção da proteína C, E, NS3hel e NS5B.....	62
<b>Figura 3.15.</b>	Resultados da purificação por cromatografia de afinidade com iões $\text{Ni}^{2+}$ das proteínas C, E, NS3hel e NS5B a partir de lisados bacterianos.....	63
<b>Figura 3.16.</b>	Detecção específica das proteínas recombinantes.....	65

# ÍNDICE DE TABELAS

## 1. MATERIAL E MÉTODOS

<b>Tabela 2.1.</b> Condições utilizadas nas amplificações, por PCR, dos fragmentos de DNA (sedimentos celulares) da região ITS do genoma das células C6/36.....	21
<b>Tabela 2.2.</b> Condições utilizadas nas amplificações, por PCR, dos fragmentos de cDNA das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do envólucro, domínio helicase da proteína NS3 e NS5B).....	22
<b>Tabela 2.3.</b> Condições utilizadas para as reacções de transcrição reversa a partir de RNA viral.....	23
<b>Tabela 2.4.</b> Informação relativa ao conjunto de <i>primers</i> utilizados nas amplificações, por PCR, de cDNA de CFAV efectuados ao longo deste trabalho.....	25
<b>Tabela 2.5.</b> Condições utilizadas nas amplificações, por PCR, dos fragmentos de DNA das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do envólucro, NS3hel e NS5B).....	30

## 2. RESULTADOS

<b>Tabela 3.1.</b> Frequência com que aparecem os códons raros nas sequências das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do envólucro, NS3hel e NS5B).....	52
---	----

## ABREVIATURAS

<b>APS</b>	Persulfato de amónia
<b>C</b>	Proteína estrutural da cápside
<b>°C</b>	Graus Celsius
<b>cDNA</b>	DNA complementar
<b>CFAV</b>	<i><u>C</u>ell <u>F</u>using <u>A</u>gent <u>V</u>irus</i>
<b>CSA</b>	<i><u>C</u>ell <u>S</u>ilent <u>A</u>gent</i>
<b>DNA</b>	Ácido desoxirribonucleico, do inglês <i><u>D</u>eoxyribo<u>n</u>ucleic <u>A</u>cid</i>
<b>dNTP</b>	Desoxirribonucleótidos 5'-trifosfatados
<b>D.O.</b>	Densidade Óptica
<b>E</b>	Glicoproteína estrutural de superfície
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	Etilenodiaminotetracetato
<b><i>et al</i></b>	Do latim, <i>et alii</i> , que significa, entre outros
<b>EUA</b>	<u>E</u> stados <u>U</u> nidos da <u>A</u> mérica
<b><i>g</i></b>	Aceleração da gravidade
<b>h</b>	Hora
<b>His</b>	Histidina
<b>IPTG</b>	Isopropil-β-D-tiogalactopiranosídeo
<b>ISFs</b>	<i><u>I</u>nsect-<u>S</u>pecific <u>F</u>laviviruses</i>
<b>kDa</b>	Kilo(Daltons) = 10 <sup>3</sup> Daltons
<b>K(p)b</b>	Kilo(pares) de bases = 10 <sup>3</sup> (pares) de bases
<b>LB</b>	Meio de cultura <u>L</u> uria <u>B</u> ertani
<b>MgCl<sub>2</sub></b>	Cloreto de magnésio
<b>NaCl</b>	Cloreto de sódio
<b>min</b>	Minuto
<b>ml</b>	Mililitro
<b>mM</b>	Milimolar
<b>mRNA</b>	Ácido ribonucleico mensageiro, do inglês <i><u>M</u>essenger <u>R</u>ibonucleic <u>A</u>cid</i>
<b>NaOH</b>	Hidróxido de sódio
<b>NCBI</b>	<i><u>N</u>ational <u>C</u>enter for <u>B</u>io<u>t</u>echnology <u>I</u>nformation</i>
<b>nm</b>	Nanómetro = 10 <sup>-9</sup> metros

<b>NS3hel</b>	Proteína não estrutural 3 (domínio com que confere à proteína actividade de helicase de RNA)
<b>NS5B</b>	Proteína não estrutural 5 (polimerase de RNA dependente de RNA)
<b>nt</b>	Nucleótido
<b>ORF</b>	<i>Open Reading Frame</i>
<b>pb</b>	Pares de base
<b>PCR</b>	Reacção de Polimerase em cadeia, do inglês <i>Polymerase Chain Reaction</i>
<b>pH</b>	Simétrico do logaritmo decimal da concentração hidrogeniónica de uma solução
<b>poli-A</b>	Cauda poliadenilada
<b>RNA</b>	Ácido ribonucleico, do inglês <i>Ribonucleic Acid</i>
<b>RNase</b>	Ribonuclease
<b>RpRd</b>	<i>RNA polimerase RNA dependente</i>
<b>RT</b>	<i>Reverse Transcription</i>
<b>RT-PCR</b>	Polimerização em cadeia da polimerase antecedita de transcrição reversa
<b>SDS</b>	Dodecilsulfato de sódio, do inglês <i>Sodium Dodecyl Sulphate</i>
<b>SDS-PAGE</b>	Electroforese em gel de Dodecilsulfato de Sódio poliacrilamida
<b>seg</b>	Segundos
<b>SOC</b>	Meio de cultura enriquecido para o crescimento bacteriano
<b>TAE</b>	Tampão Tris-acetato EDTA
<b>TE</b>	Tampão Tris-EDTA
<b>TEMED</b>	N, N, N', N' - tetrametil-etileno-diamina
<b>Tris</b>	Tris Hidrocloreto
<b>Tris-HCL</b>	Tris(hidroximetil)aminometano
<b>TSS</b>	<i>Transformation and Storage Solution</i>
<b>UTR</b>	Região não traduzida, do inglês, <i>Untranslated Region</i>
<b>UV</b>	Ultravioleta
<b>vs</b>	<i>versus</i>
<b>μ</b>	Micro, prefixo que designa 10 <sup>-6</sup>
<b>μl</b>	Microlitros



# ***1. INTRODUÇÃO***

---

## 1. INTRODUÇÃO

### 1.1. Considerações gerais

A família *Flaviviridae* é organizada em três géneros bem estabelecidos: *Hepacivirus*, que inclui o vírus da Hepatite C (HCV-*Hepatitis C Virus*), o hepacivírus canino e, provavelmente, o vírus GBV-B; *Pestivirus*, cujo principal representante é o vírus da diarreia viral bovina (BVDV-*Bovine Viral Diarrhea Virus*) e *Flavivirus* (Wu *et al.*, 2005; Lindenbach *et al.*, 2007; Kapoor *et al.*, 2011). Para além destes três géneros, recentemente foi proposta a inclusão na família *Flaviviridae* de um quarto género, denominado de *Pegivirus*, que deverá integrar os vírus GBV-A, GBV-C e GBV-D (Stapleton *et al.*, 2011).

O género *Flavivirus* é constituído por um grande número de vírus (mais de 70), alguns dos quais são importantes agentes patogénicos para o homem e para outros animais, com o potencial para provocarem grandes epidemias associadas a elevadas taxas de mortalidade e morbilidade, e dos quais se destacam, entre muitos outros, o vírus da Dengue (DENV), o vírus da Encefalite Japonesa (JEV), o vírus do Nilo Ocidental (WNV) e o vírus da Febre Amarela (YFV) (Gritsun e Gould, 2006; Lindenbach *et al.*, 2007; Perera *et al.*, 2008; Bollati *et al.*, 2010). Este género está subdividido em quatro grupos principais, nos quais os agentes virais são divididos com base nas suas relações filogenéticas, sendo que esta divisão é apoiada, em grande medida, pelas diferentes características ecológicas dos respectivos hospedeiros. Assim, num destes grupos encontramos vírus transmitidos por mosquitos, noutros vírus transmitidos por carraças, e num terceiro inclui a maioria dos flavivírus zoonóticos sem vector invertebrado conhecido (NKV-*No Known Vector*) (Hoshino *et al.*, 2007). No entanto, e sobretudo durante os últimos anos, inúmeros estudos têm vindo a colocar em evidência o facto de que esta classificação não esgota a diversidade genética dos flavivírus.

É disso exemplo a descoberta do *Tamara Bat Virus* (TABV), que apesar de ter sido isolado a mais de três décadas e de estar mais estreitamente relacionado com os flavivírus, do que com outros géneros (*Pestivirus* e *Hepacivirus*), ainda não foi determinada sua posição filogenética. Recentemente, a determinação completa da sua sequência genómica veio apoiar a sugestão de que este vírus venha a integrar um género diferente dentro da família *Flaviviridae* (de Lamballerie *et al.*, 2002).

À parte do TABV, um quarto grupo inclui já vários vírus conhecidos, tais como o *Cell Fusing Agent Virus* (CFAV), o *Kamiti River Virus* (KRV), o *Culex Flavivirus* (CxFV) ou o *Quang Binh Virus* (QBV). A análise de um grande número de populações naturais de mosquitos indica que estes vírus formam um grupo geneticamente diverso, com uma maior prevalência e dispersão geográfica do que inicialmente se antecipou. Ainda que filogeneticamente distintos da grande maioria dos demais flavivírus, a sua inserção no género é sustentada por semelhanças com flavivírus “clássicos” em termos de morfologia dos viriões, organização dos seus genomas, e de identidade de sequências em certos componentes estruturais e funcionais (para além dos perfis de hidrofília) das diferentes proteínas virais. No entanto, ao contrário da maioria dos flavivírus, eles não replicam, pelo menos *in vitro*, em células de vertebrados, justificando a sua designação como “flavivírus específicos de insectos” (ISF- *insect-specific flavivirus*) (Crabtree *et al.*, 2003, Sang *et al.*, 2003; Hoshino *et al.*, 2007; Blitvich *et al.*, 2009; Cook *et al.*, 2006, 2009, 2012; Parreira *et al.*, 2012). Neste grupo, por sua vez, os vírus dividem-se, pelo menos, em três outras linhas distintas, uma constituída por vírus associados a mosquitos do género *Culex*, outra por vírus associados a mosquitos do género *Aedes* (Crabtree *et al.*, 2003, Sang *et al.*, 2003; Cook *et al.*, 2006) e uma terceira por vírus associados a mosquitos do género *Ochlerotatus* (Huhtamo *et al.*, 2012)

Embora o número de estudos envolvendo a identificação e caracterização de ISF tenha sofrido uma “explosão” em anos recentes, o primeiro destes vírus (CFAV) foi identificado há mais de 3 décadas (ver secção seguinte). Recentemente, o CFAV foi encontrado em outros culicídeos de diversas áreas do mundo, particularmente nos membros do género *Aedes* (*Aedes aegypti*, *Aedes albopictus*) e, curiosa e inesperadamente, também entre alguns mosquitos do género *Culex*. São destes estudos exemplo os que reportam a identificação de CFAV em populações naturais de mosquitos colhidos em Porto Rico (Cook *et al.*, 2006), ou a identificação de sequências específicas do CFAV em homogeneizados de mosquitos *Aedes aegypti* colhidos na Tailândia por Kihara e colaboradores (2007).

Estudos com base em inferência filogenética, demonstraram que o CFAV e KRV ocupam uma linhagem mais divergente dentro da linhagem ISF, sugerindo que são os flavivírus primordiais e que, provavelmente, emergiram antes dos outros membros deste género a partir de um ancestral comum (Crabtree *et al.*, 2003; Peyrefitte *et al.*, 2007;

Hoshino *et al.*, 2007; Kim *et al.*, 2009), sendo ainda possível que os vírus transmitidos por artrópodes possam ter evoluído a partir de vírus específicos de insectos (Cook *et al.*, 2006).

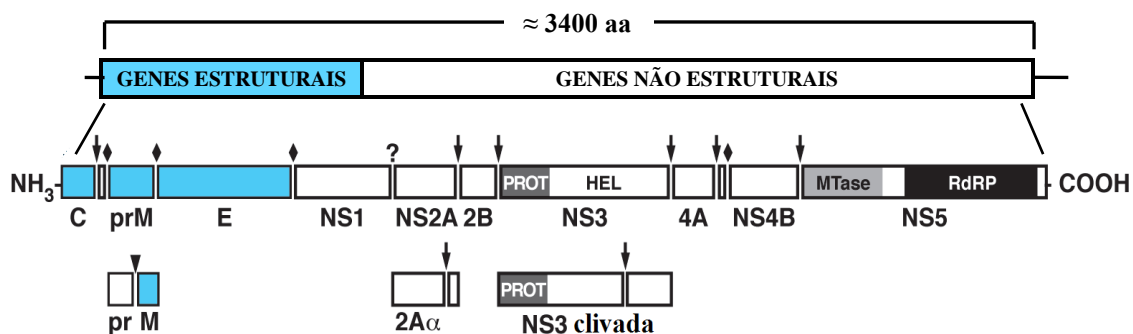
Talvez, o mais surpreendente de tudo foi a identificação de sequências de DNA filogeneticamente relacionada com os genomas do CFAV e KRV, e identificadas no genoma de células de *Aedes albopictus* (C6/36), assim como no quer de mosquitos selvagens e criados no laboratório em diversas áreas do mundo (Crochu *et al.*, 2004). No caso específico das células de *Aedes albopictus* (C6/36), tais sequências deverão representar a integração de sequências de um vírus denominado *Cell Silent Agent* (CSA; hoje em dia claramente defectivo em múltiplas ocasiões). Estes resultados podem indicar que os ISF estão intimamente relacionados com os mosquitos e que podem existir em ambientes naturais que não foram ainda identificados até a data (Cook *et al.*, 2006; Hoshino *et al.*, 2007). Tal facto, só por si, levantou questões importantes tais como possíveis implicações importantes no que diz respeito a evolução (uma vez que representa um método para transferência horizontal de genes em células eucarióticas; Kim *et al.*, 2009), dos efeitos de mortalidade e morbilidade dos ISF em mosquitos e suas interações com estirpes patogénicas ser completamente desconhecidas (Parreira *et al.*, 2012).

## **1.2. *Cell Fusing Agent Virus* (CFAV): da sua descoberta à actualidade**

Há mais de três décadas, Stollar e Thomas (1975), isolaram no laboratório um agente viral, o qual veio a ser designado de *Cell Fusing Agent Virus* (CFAV), a partir de uma linha de células de *Aedes albopictus* (C6/36) usando um macerado de mosquito *Aedes aegypti*. Este nome foi-lhe dado em virtude do característico efeito citopático aquando da sua replicação nestas células, o qual se traduz pela formação de sincícios enormes.

Os viriões do CFAV apresentam uma morfologia esférica e dimensões aproximadas de 50 nm de diâmetro. Possuem um invólucro lipídico e um genoma de RNA em cadeia simples e polaridade positiva [ssRNA (+)]. Cerca de 15 anos mais tarde, a sequenciação do genoma completo deste vírus veio revelar uma única grelha de leitura aberta (ORF- *Open Reading Frame*) a qual codifica uma única poliproteína (Cammisa-Parks *et al.*, 1992). Esta ORF é flanqueada por regiões não codificantes

(UTRs – *Untranslated Terminal Regions*) nas extremidades 5' (estabilizado por uma estrutura *cap*) e 3', a qual é desprovida de cauda de poli (A) (Gritsun e Gould, 2006; Lindenbach *et al.*, 2007). No decurso da replicação viral, e à semelhança do que ocorre durante a replicação dos flavivírus que foram, até hoje, estudados, esta poliproteína (com aproximadamente 3400 resíduos de aminoácidos) deverá ser processada pela actividade combinada de proteases virais e celulares, dando origem a 3 proteínas ditas estruturais (proteína da cápside, proteína da membrana e glicoproteína do invólucro) e 7 proteínas não-estruturais (NS1, NS2A/B, NS3, NS4A/B e NS5), estas últimas desempenhando funções regulatórias e enzimáticas. A organização do genoma dos flavivírus encontra-se representada na figura 1.1.



**Figura 1.1.** Representação esquemática da organização do genoma e processamento da poliproteína dos flavivírus. As caixas abaixo do genoma, onde estão indicadas as regiões que codificam os componentes estruturais e não estruturais, indicam precursores e proteínas maduras originadas pela cascata de processamento proteolítico. As proteínas estruturais estão coloridas de azul e as proteínas não estruturais (NS) estão coloridas de branco. São indicados os locais de clivagem: (♦) proteases celulares, (↓) proteases viral, (▼) uma furina ou protease, e (?) proteases desconhecidas (Adaptado de Lindenbach *et al.*, 2007).

Estas características correspondem a algumas das que apoiam a inclusão do CFAV no género *Flavivirus* da família *Flaviviridae* (Cook *et al.*, 2006; Peyrefitte *et al.*, 2007; Bollati *et al.*, 2010). Apesar de, após o seu isolamento, o CFAV ter sido classificado juntamente com os demais vírus dos géneros *Flavivirus* e *Pestivirus*, como sendo um possível membro da família *Togaviridae* (Westaway *et al.*, 1985), a sua inclusão no

género *Flavivirus* da família *Flaviviridae* foi homologada pelo Comité Internacional de Taxonomia de Vírus (Peyrefitte *et al.*, 2007).

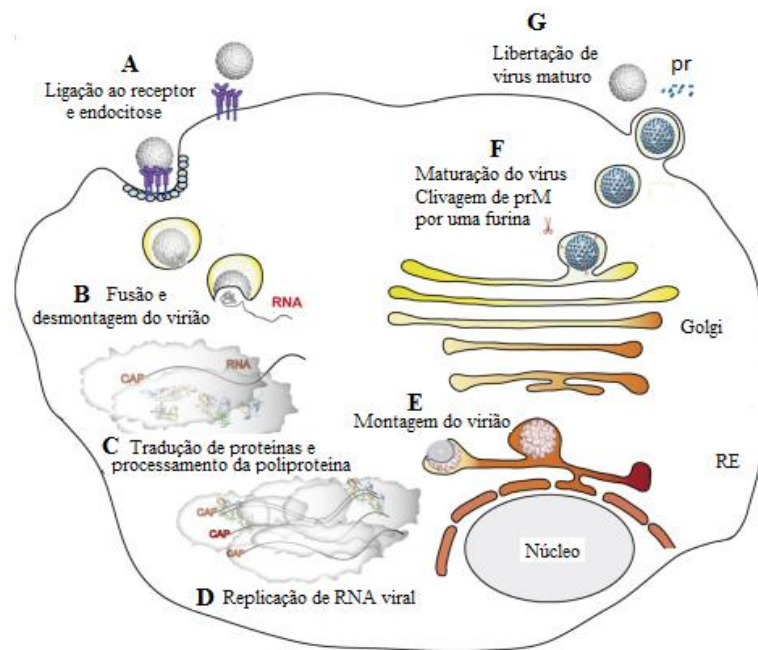
Mas apesar de ser considerado como sendo o primeiro ISF (Crabtree *et al.*, 2003, Sang *et al.*, 2003; Hoshino *et al.*, 2007; Blitvich *et al.*, 2009; Cook *et al.*, 2009), a biologia do CFAV, como, aliás, a da totalidade dos demais ISF, não foi, até hoje, minimamente caracterizada. Nada se sabe sobre a sua replicação *in vivo* ou *in vitro*, sendo que a maioria dos estudos até hoje realizados abordam os ISF de uma forma descritiva, centrando-se, essencialmente, sobre a sua análise do ponto de vista filogenético. Por analogia, é expectável que o seu ciclo replicativo reproduza as características do descrito para os demais flavivírus (essencialmente patogénicos) estudados até hoje.

### 1.3. Ciclo replicativo dos flavivírus

Os flavivírus replicam no citoplasma das células que infectam (ver Fig. 1.2), sendo que o ciclo replicativo é iniciado com a ligação dos vírus à superfície da célula por múltiplas interacções entre receptores virais e celulares. As partículas virais são internalizadas por endocitose. O pH baixo resultante da acidificação dos endossomas primários provoca o rearranjo estrutural das glicoproteínas virais que promove a fusão da partícula viral com a membrana do endossoma, a descapsidação do virião e a consequente libertação do RNA viral no citoplasma.

A replicação do genoma viral ocorre em complexos replicativos citoplasmáticos associados a membranas na região perinuclear, e depende da síntese de um RNA intermediário de polaridade negativa (-) formando um intermediário de cadeia dupla (dsRNA) que, por sua vez, será usado para a síntese de novas moléculas de RNA de polaridade positiva (+). Estas servirão quer como mRNA quer como RNA genómico, a ser encapsidado nas novas partículas virais (Lindenbach *et al.*, 2007; Perera *et al.*, 2008). A morfogénese é coincidente com a encapsidação de moléculas de RNA genómico nas novas partículas virais e ocorre ao nível da membrana do retículo endoplasmático (RE). Nestas, as proteínas da cápside, o RNA genómico e as glicoproteínas são envolvidos pela membrana do RE, formando partículas virais imaturas. Estas são transportadas para região distal do aparelho de Golgi (*trans*-Golgi)

onde sofrem a maturação, e em seguida são encaminhadas para o exterior de células em vesículas intracelulares, da qual são libertadas por exocitose (Lindenbach *et al.*, 2007; Perera *et al.*, 2008; Sampath e Padmanabhan, 2009).



**Figura 1.2.** Ciclo replicativo dos flavivírus (Adaptado de Perera *et al.* 2008).

#### 1.4. Proteínas estruturais e não estruturais dos flavivírus

A proteína da cápside (C) é a primeira proteína estrutural a ser traduzida. É uma proteína altamente básica, com massa molecular de aproximadamente de 11kDa, e que interage com o RNA genómico viral, formando uma nucleocápside (Sampath e Padmanabhan, 2009). Possui um domínio hidrofóbico no extremo carboxilo, através do qual interage com as membranas celulares, e desempenha um papel importante na encapsidação do genoma viral nas partículas virais em formação (Kofler *et al.*, 2002; Jones *et al.*, 2003; Lindenbach *et al.*, 2007). No entanto, pouco se sabe sobre o reconhecimento do RNA viral pela proteína C. Para a maioria dos flavivírus, ainda não é claro como, quando, e onde a proteína C recruta o RNA viral durante a morfogénese da partícula viral (Samsa *et al.*, 2009).

A proteína estrutural M encontra-se ancorada no invólucro de partículas virais maduras e é formada a partir de uma glicoproteína precursora de membrana (prM), cujo processamento também origina o segmento pr. A importância da proteína M prende-se

com o papel que desempenha na montagem das partículas do vírus. Para além disso, protege a proteína E prematura durante as mudanças conformacionais da maturação e libertação de partículas virais (Murray *et al.*, 2008; Perera *et al.*, 2008).

A glicoproteína E, com aproximadamente 53kDa, desempenha um papel essencial no ciclo replicativo viral. É responsável por mediar a adsorção da partícula viral à célula hospedeira através da ligação a receptores celulares, e pela fusão do invólucro viral com a membrana celular, permitindo assim a libertação da nucleocápside no citoplasma da célula (Kofler *et al.*, 2002; Lindenbach *et al.*, 2007; Perera *et al.*, 2008).

Recentes avanços na determinação da estrutura dos flavivírus e das suas proteínas dão especial destaque às mudanças conformacionais sofridas pela proteína E, como as transições entre as formas imaturas, maduras, e do desempenho de um papel importante durante a montagem do vírus. Superfícies específicas da proteína E são descritas como potenciais alvos para o desenho de drogas antivirais e estratégias alternativas para a inibição viral que são abordadas com base na interacção da proteína E com moléculas receptoras e da proteína E com anticorpos neutralizantes. Além disso, a informação estrutural adquirida a partir dessas interacções podem ser utilizadas para conceber novos inibidores de entrada de vírus com eficácia aumentada (Perera *et al.*, 2008).

Por seu lado, o grupo das proteínas não estruturais engloba 7 componentes, dos quais o primeiro a ser traduzido é designado NS1.

Esta proteína, com aproximadamente 46kDa, é uma glicoproteína localizada dentro de células infectadas, bem como na sua superfície, sendo ainda eficientemente secretada pelas células infectadas por flavivírus. A proteína NS1 é clivada na junção NS1/NS2A por uma enzima provavelmente celular mas, até hoje, desconhecida. Ela actua como um cofactor da replicação de RNA viral, apesar de ainda serem desconhecidos os mecanismos moleculares envolvidos (Lindenbach e Rice, 1999; Lindenbach *et al.*, 2007; Youn *et al.*, 2010). Uma mutação na NS1 do YFV mostrou ser prejudicial na acumulação de RNA (Youn *et al.*, 2010). Lindenbach e Rice (1999) revelaram a interacção genética entre a NS1 e a NS4A, e a importância desta na replicação de RNA.



A NS2A é uma proteína hidrofóbica e relativamente pequena, apresentando uma massa molecular estimada de aproximadamente 22kDa (Lindenbach *et al.*, 2007; Geiss *et al.*, 2009), sendo um componente essencial do complexo de replicação viral (replicase). O processamento da região NS2A-NS2B-NS3 é realizado pela protease NS3 associada ao seu cofactor NS2B (ver abaixo). A NS2A também é clivada pela protease viral originando assim uma proteína na forma truncada que é designada de NS2A $\alpha$  (Murray *et al.*, 2008). Mutações no local de clivagem da NS2A $\alpha$  do vírus da Febre Amarela (YFV) mostraram evidências de que a NS2A $\alpha$  está envolvida na montagem do vírus (Lindenbach *et al.*, 2007). Além disso, a NS2A parece interagir com outras proteínas não estruturais, incluindo a NS3 (Murray *et al.*, 2008) e a NS5, e com a região UTR 3'-terminal (Lindenbach *et al.*, 2007).

A NS2B, com aproximadamente 14kDa, é uma pequena proteína associada a membrana. Esta forma um complexo estável com a proteína NS3 e actua como cofactor (NS2B-NS3 protease) (Lindenbach *et al.*, 2007; Bollati *et al.*, 2010).

Esta última (NS3), com aproximadamente 70kDa, é uma enzima multifuncional com actividades de protease, helicase (RNAhel) e trifosfatase (RTPase) de RNA (Nomaguchi *et al.*, 2004; Lindenbach *et al.*, 2007; Borowski *et al.*, 2008; Bollati *et al.*, 2010). O domínio responsável pela actividade proteásica da NS3 encontra-se localizado na sua região N-terminal, enquanto que os domínios que conferem a NS3 actividade de trifosfatase e helicase de RNA, se encontram na região C-terminal (Lindenbach *et al.*, 2007; Borowski *et al.*, 2008). Tal como acima mencionado, a protease NS3 e o seu cofactor, a proteína NS2B, estão envolvidos no processamento da poliproteína viral. Participam na clivagem das junções de NS2A-NS2B, NS2B-NS3, NS3-NS4A e NS4B-NS5 (Nomaguchi *et al.*, 2004; Bollati *et al.*, 2010). A actividade de trifosfatase de RNA está envolvida na síntese da estrutura 5'-cap do genoma viral (Crabtree *et al.*, 2003; Nomaguchi *et al.*, 2004; Bollati *et al.*, 2010).

O domínio proteico que confere a NS3 actividades de helicase de RNA está envolvido na separação de cadeias de RNA formadas durante a replicação viral e na remoção de proteínas ligadas ao RNA viral (Borowski *et al.*, 2008; Geiss *et al.*, 2009). A separação de cadeias é uma reacção dependente de energia dirigida pela hidrólise de ATP. Portanto, todas as helicases de RNA possuem, ainda, actividade de ATPase. No

entanto, as helicases de RNAs codificadas pelos flavivírus podem hidrolisar diferentes nucleósidos trifosfatados, razão pela qual esta actividade enzimática é vulgarmente descrita como NTPase (Sampath e Padmanabhan, 2009). Adicionalmente, a proteína NS3 através da sua actividade de helicase de RNA, participa no início da replicação, promovendo o “desenrolamento” de estruturas secundárias que tendem a formar-se especialmente na região não traduzida da extremidade 3' do RNA viral (Wu *et al.*, 2005).

A estrutura da helicase de RNA dos flavivírus consiste em três subdomínios, dois dos quais são conservados entre os membros desta família de proteínas, e estão envolvidos na hidrólise de nucleosídeos trifosfatados, e um subdomínio C-terminal que pode estar envolvido no reconhecimento de proteínas e RNA de vírus específicos (Wu *et al.*, 2005; Lindenbach *et al.*, 2007). Análise mutacional mostrou que alterações das actividades de RNAhel, RTPase ou ATPase resultam na redução ou supressão total da replicação viral (Geiss *et al.*, 2009).

Devido às suas características enquanto enzima e ao seu papel crucial no ciclo replicativo viral, quer através na sua participação como componente da replicase (ver abaixo), quer como a principal protease implicada no processamento da poliproteína codificada pelos genomas dos flavivírus, NS3 constitui um dos alvos prioritários para o desenvolvimento de novas drogas com actividade antiviral (Sampath *et al.*, 2006; Borowski *et al.*, 2008; Assenberg *et al.*, 2009; Bollati *et al.*, 2010).

A NS4A e NS4B são pequenas proteínas (16kDa e 27kDa, respectivamente) que através das suas inúmeras regiões hidrofóbicas permanecem associadas às membranas intracelulares (Lindenbach *et al.*, 2007). As suas funções ainda são pouco conhecidas, mas estas proteínas parecem ter um papel importante na replicação viral. Supõe-se que desempenham um papel na indução de alterações de membrana, adoptadas para servirem de suporte para a formação do complexo de replicação viral (Roosendaal *et al.*, 2006).

Finalmente, a proteína NS5 é a maior (com aproximadamente 100kDa) e mais conservada das proteínas virais (Tan *et al.*, 1996). Parece possuir duas actividades enzimáticas que desempenham um papel fulcral na replicação viral. São por elas responsáveis um domínio localizado na região N-terminal e que lhe confere actividade

de metiltransferase (NS5MTase) e que justifica o seu envolvimento na formação da estrutura *cap* presente na região 5'-terminal do RNA genómico viral, enquanto que na região C-terminal um grande domínio com vários motivos conferem a proteína NS5 actividade de polimerase de RNA dependente de RNA (RpRd) (Tan, *et al.*, 1996; Nomaguchi *et al.*, 2004; Geiss *et al.*, 2009; Bollati *et al.*, 2010). Assim sendo, a proteína NS5 é a principal proteína responsável pela síntese de cadeia de RNA (-) e cadeia de RNA genómico (+) (Geiss *et al.*, 2009). Vários estudos que permitiram demonstrar que proteínas NS5 codificadas por alguns flavivírus [tais como, o vírus da Dengue (DENV), vírus Kunjin (KUNV), vírus da Encefalite Japonesa (JEV)] e pelo vírus da Hepatite (HCV), expressas como proteínas recombinantes apresentavam actividade de RpRd, sendo capazes de iniciar a replicação de RNA viral na ausência de outros cofactores virais e celulares (Tan *et al.*, 1996; Ferrari *et al.*, 1999; Guyatt *et al.*, 2001; Nomaguchi *et al.*, 2004; Yu *et al.*, 2007; Murray *et al.*, 2008; Bollati *et al.*, 2010). Há evidência de que durante o processo de replicação a proteína NS5 interacciona com o domínio C-terminal da proteína NS3, isto é, a NS5RpRd reconhece e liga-se a dsRNA (+), a partir das quais, e em combinação com actividade da proteína NS3helicase sintetizam uma nova cadeia de RNA (+). As actividades da NS5MTase e da NS3RTPase auxiliam a formação da estrutura *cap* do RNA viral (Assenberg *et al.*, 2009; Geiss *et al.*, 2009). À semelhança do que acontece com a proteína NS3, a proteína NS5 tem sido alvo de inúmeros estudos que visam o desenvolvimento inibidores com actividade de baixa toxicidade e actividade *in vivo*, os quais possam poder vir a ser utilizados como novas drogas para a terapia antiviral.

### 1.5. Sistemas de expressão de proteínas recombinantes

Ao longo dos últimos anos têm sido desenvolvidos processos de produção de proteínas com várias aplicações, nomeadamente, em medicina (quer em diagnóstico quer em terapêutica), na agricultura (plantas modificadas), na obtenção de produtos alimentares e na tecnologia associada, por exemplo, ao uso de enzimas (Demain e Vaishnav, 2009).

A escolha de um sistema de expressão para a elevada produção de proteínas recombinantes depende de alguns factores. Estes incluem as características e requisitos

nutricionais que garantem o crescimento celular dos hospedeiros de expressão, o nível de expressão proteica pretendido, a localização celular das proteínas a produzir (expressão intracelular ou extracelular), a sua eventual modificação pós-tradução, e a manutenção da sua actividade biológica hipotética. Para além disso, a selecção de um sistema de expressão particular exige uma repartição dos custos em termos de processo, construção, e outras considerações de ordem económica (Makrides, 1996).

São duas as principais abordagens genéricas para a expressão heteróloga de proteínas recombinantes: o sistema “procariota” (baseado na utilização de células procariontes como as bactérias) e o sistema “eucariota” (fazendo o uso de células eucariontes tais como leveduras, células de insectos, plantas e mamíferos) (Demain e Vaishnav, 2009).

Os sistemas de expressão baseados na utilização de hospedeiros procariontes são, geralmente, a primeira escolha para a produção de proteínas recombinantes. A tal é devida o facto de as bactérias normalmente utilizadas (na grande maioria das vezes *Escherichia coli*) são fáceis de manipular, apresentam um crescimento rápido em meios de composição bem conhecida e pouco dispendiosa, e permitem expressão proteica de alta produtividade a baixo custo (Terpe, 2006). São, sobretudo, excelentes sistemas para a expressão de proteínas funcionais não glicosiladas. No entanto, apresentam algumas desvantagens, nomeadamente, o facto de que a sobre-expressão de proteínas no citoplasma das células bacterianas levar, frequentemente, à formação de agregados insolúveis de proteínas (designadas corpos de inclusão). Por outro lado, a elevada densidade celular leva à acumulação de acetato resultando em toxicidade celular, para além de que são genericamente incapazes de introduzir modificações pós-traducionais nas proteínas que expressam (Terpe, 2006; Demain e Vaishnav, 2009).

Assim sendo, a escolha de sistemas que dependam da utilização de células eucariontes surge como alternativa à expressão em bactérias, especialmente quando são necessárias modificações pós-traducionais, frequentemente essenciais quando se pretende, por exemplo, a utilização das proteínas expressas em ensaios funcionais para caracterização da sua função biológica e/ou das interacções que estabelecem com outras proteínas. No entanto, este sistema apresenta algumas deficiências, como por exemplo, o sistema de células de insectos é inferior aos sistemas “procariotas” e leveduras em termos de sua capacidade de fermentação contínua. Além disso, as células de insectos e as células de mamíferos diferem nos seus padrões de glicosilação (Yin *et al.*, 2007;

Demain e Vaishnav, 2009), como nos comprimentos dos oligossacarídeos e no seu conteúdo em manose (Yin *et al.*, 2007). Por outro lado, o sistema de culturas de células de mamíferos apresenta outras limitações como o custo relativamente alto, tecnologia complexa, potencial contaminação com vírus e o uso destes sistemas requerem um período de produção relativamente mais longo, (Yin *et al.*, 2007; Demain e Vaishnav, 2009).

### **1.5.1. *Escherichia coli* como hospedeiro de expressão de proteínas recombinantes**

A bactéria Gram-negativa *E. coli* é o organismo mais comumente usado para a produção de proteínas heterólogas (Terpe, 2006). Uma das razões para tal prende-se com o facto deste organismo, em termos genéticos e da sua biologia molecular, ser bem caracterizado (Makrides, 1996; Baneyx, 1999). A expressão de proteínas baseada na utilização de *E. coli* apresenta inúmeras vantagens. De entre elas destacam-se o seu rápido crescimento, fácil sub-cultura e conservação, e a rápida expressão das proteínas a produzir, as quais podem ser obtidas com, elevado rendimento e a custo relativamente baixo (Yin *et al.*, 2007; Sahdev *et al.*, 2008; Demain e Vaishnav, 2009). Obviamente, a produção de proteínas heterólogas em *E. coli* apresenta várias limitações. De entre estas destaca-se o facto de que, muitas proteínas são complexas, apresentando subunidades múltiplas, cofactores (Mahmoud, 2007), ou modificações pós-traducionais (como por exemplo, N- e O- glicosilação, acilação dos ácidos gordos, fosforilação, e estabelecimento de ligação dissulfeto) as quais são essenciais para a sua função (Yin *et al.*, 2007). Nestes casos, a sua produção usando *E. coli* não é aconselhada.

Embora *E. coli* tenha uma notável capacidade para produzir proteínas com elevado rendimento, muitas vezes estas são expressadas em quantidades limitadas, especialmente se se tratarem de proteínas com origem em células eucariontes ou codificadas pelos vírus que as infectam. Uma das razões para tal prende-se com o desajuste que pode existir entre o uso de codões que codificam a proteína de interesse relativamente aos de *E. coli* (Makrides, 1996; Sahdev *et al.*, 2008). Para além deste, se o mRNA de interesse contém muitos codões, que são raros em *E. coli*, a diminuição da quantidade de tRNAs daí decorrente pode, ainda, levar à incorporação de aminoácidos

errôneos e/ou o truncamento do polipéptido a ser sintetizado, causando efeitos sobre os níveis de expressão e a qualidade de proteínas heterólogas produzidas (Mirzahoseini *et al.*, 2011). Duas estratégias alternativas são, então, utilizadas para corrigir o problema da presença de codões raros. Por um lado podemos socorrer-nos de uma abordagem baseada na mutagénesse dirigida da sequência codificante da proteína de interesse, de forma a gerar codões que sejam reconhecidos pelo “pool” de tRNAs do sistema hospedeiro, sem modificar a sequência primária da proteína codificada. Por outro, fazer uso da co-expressão no hospedeiro procarionte dos tRNAs que reconheçam codões raramente utilizados por este, normalmente a partir de plasmídeos (Hannig e Makrides, 1998; Sahdev *et al.*, 2008; Mirzahoseini *et al.*, 2011).

Após síntese da proteína de interesse, esta necessita de adquirir uma estrutura tridimensional específica, da qual depende directamente a sua função, por um processo designado de *folding* proteico (Bross e Gregersen, 2003; Vabulas *et al.*, 2010). Alguns dos autores traduzem este processo para português por “enrolamento”, mas neste trabalho optou-se por manter a designação anglo-saxónica. O processo de *folding* proteico tem sido um assunto intensamente estudado desde os trabalhos pioneiros de Anfinsen e seus colaboradores acerca do *folding/unfolding* da ribonuclease A (Anfinsen, 1973). Estes mostraram que uma cadeia polipeptídica pode adquirir o *folding* espontaneamente *in vitro* sem requerer a assistência de nenhum factor celular (Anfinsen, 1973), provando que toda a informação necessária para uma proteína adquirir a sua conformação nativa se encontra codificada na sua sequência linear de aminoácidos. O processo de *folding* inicia-se com a síntese da cadeia polipeptídica e termina com a conformação nativa da proteína. No entanto, o processo de *folding* ganha em complexidade quando a proteína obtida possui resíduos de cisteínas que estabelecem ligações dissulfeto. A presença de ligações dissulfeto é importante na manutenção do processo de *folding* de muitas proteínas secretadas para o exterior do citoplasma. Em contraste, proteínas encontradas no citoplasma estabelecem ligações dissulfetos apenas como parte de seu ciclo catalítico e não como uma forma de estabilização do processo de *folding* (Hiniker e Bardwell, 2003). No entanto, a *E. coli* tem capacidade limitada de realizar ligações dissulfeto, do qual pode resultar na obtenção de proteína inactiva. Esta perda de actividade biológica da proteína produzida, também, pode ser devido a formação de agregados insolúveis (designados de corpos de inclusão) (Sahdev *et al.*, 2008).

De forma a lidarem com os desafios do processo de *folding*, os sistemas biológicos, nomeadamente os procariontes, desenvolveram uma maquinaria proteica específica - o sistema de controlo de qualidade proteica. Este compreende os chaperões moleculares, proteases e factores reguladores que regulam a actividade ou permitem a comunicação entre os vários componentes (Gregersen *et al.*, 2006; Sahdev *et al.*, 2008). Este sistema tem a capacidade de supervisionar o *folding*, contrariar a agregação e seleccionar e eliminar polipéptidos com baixa capacidade de *folding* ou com erros que poderiam exercer efeitos tóxicos para a célula (Bross e Gregersen, 2003; Hartl e Hayer-Hartl, 2009; Vabulas *et al.*, 2010). Este sistema é também capaz de se adaptar a stress ambientais como alterações de pH, temperatura e stress oxidativo, os quais afectam fortemente o processo de expressão de proteína recombinante (Bross e Gregersen, 2003; Sahdev *et al.*, 2008).

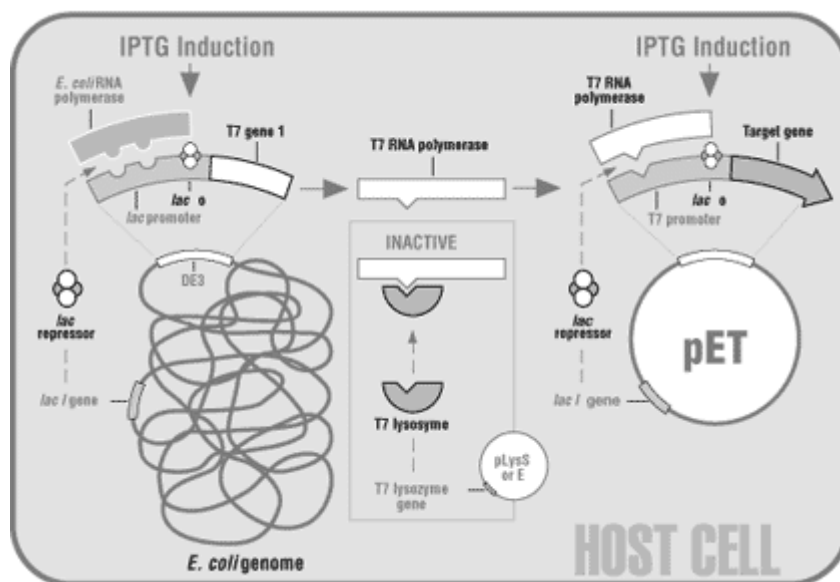
Para a produção da proteína heteróloga de interesse são vulgarmente utilizados três tipos de vectores de expressão: plasmídeos, bacteriófagos e cosmídeos (Snyder e Champness, 1997). Os vectores de expressão são capazes de auto-replicar e regular a expressão dos genes neles codificados.

A expressão proteica é ainda dependente das elevadas taxas de transcrição conseguidas através da inclusão nos vectores de promotores de transcrição forte e facilmente indutíveis, os quais incluem, os frequentes promotores de transcrição T7-*lac*, *lac*, *tac*, *trc*, (Snyder e Champness, 1997; Sorensen e Mortensen, 2005), ao mesmo tempo, firmemente regulados de forma a garantir um nível de expressão basal (ou residual) baixa, especialmente nos casos em que a proteína de interesse é tóxica ou, de alguma forma prejudica o crescimento da célula hospedeira (Hannig e Makrides, 1998; Demain e Vaishnav, 2009).

### 1.6. O sistema pET

O sistema de expressão comercialmente conhecido por pET (Novagen) é dos mais frequentemente usado quando se pretende produzir uma proteína heteróloga usando *E. coli* como hospedeira de expressão, e inclui uma panóplia de vectores plasmídicos e estirpes bacterianas. Os vectores pET integram uma família de vectores a partir dos quais a expressão proteica pode ser induzida e controlada através da acção

combinada da polimerase de RNA do fago T7 e dos promotores de transcrição por esta reconhecidas (Snyder e Champness, 1997; Baneyx, 1999). A estratégia geral para a utilização de um vector pET é ilustrada na figura 1.3.



**Figura 1.3.** Estratégia para a regular a expressão de genes clonados dentro de um vector pET (Adaptado do catálogo 2009/2010, pág 79, Novagen, Madison, EUA).

A indução da expressão das proteínas de interesse é promovida pela activação da transcrição dos genes que as codificam pela polimerase de RNA do fago T7 presente nas células hospedeiras. Esta enzima é extremamente activa e selectiva, fazendo com que parte considerável dos recursos biológicos da bactéria hospedeira sejam mobilizados para a expressão dos genes pretendidos.

A expressão proteica ocorre, por seu lado, numa estirpe de expressão particular que apresenta no seu cromossoma um gene que codifica uma polimerase de RNA do fago T7 (sob o controlo transcrricional do promotor *lacUV5*) integrada no genoma de uma variante do fago lambda (DE3), o qual pode ser induzido com IPTG (Isopropil- $\beta$ -D-tiogalactopiranosídeo). Esta estirpe de expressão possui, ainda, um plasmídeo designado pLysS (ou um dos seus derivados) que transporta um gene que codifica a lisozima do fago T7, a qual é, por sua vez, inibidora natural da respectiva polimerase de RNA e, por consequência, permite manter o nível de expressão proteica basal a um nível reduzido. Por outro lado, os vectores de expressão possuem um operador *lac* a jusante do promotor T7, para além do que transportam o promotor natural e a sequência que



codifica para o repressor *lac* (*lacI*). Este repressor *lac* actua tanto no promotor *lacUV5* no cromossoma da célula hospedeira, ao reprimir a transcrição do gene da polimerase de RNA do fago T7 feita pela polimerase do hospedeiro, como no promotor *T7lac* no vector, bloqueando a transcrição do gene alvo feita pela polimerase de RNA do fago T7 expressa. Desta forma, este sistema oferece um controlo bastante apertado da expressão basal dos genes clonados, reduzindo o risco de perder clones recombinantes, por toxicidade do produto de expressão para a célula hospedeira (Mierendorf *et al.*, 1999).

### 1.7. Notas finais e objectivos

Apesar de o CFAV ter sido identificado há mais de três décadas (1975), a verdade é que muito pouco se conhece sobre a sua biologia. Dado o seu parentesco filogenético estreito com alguns outros flavivírus encontrados naturalmente em mosquitos de diferentes géneros colhidos em diferentes regiões do globo, este vírus poderá ser usado como modelo para o estudo dos ISF.

A análise da estratégia replicativa utilizada pelo CFAV e as interacções que estabelece, *in vitro* e *in vivo*, com as células em que replicam carecem, no entanto, do desenvolvimento de ferramentas básicas, tais como clones moleculares ou baterias de soros contendo anticorpos que reconheçam uma ou mais proteínas codificadas pelo genoma viral. Assim sendo, foi nosso intuito obter, de forma recombinante, 4 das proteínas codificadas pelo CFAV [duas estruturais (C e E) e duas não estruturais (NS3hel e NS5B)], as quais poderão vir a ser utilizadas tanto como antigénios para produção de anticorpos, cristalizadas para determinação da sua estrutura tridimensional e/ou, e em especial no que diz respeito a NS3hel e NS5B, directamente utilizadas em ensaios que visem a sua caracterização bioquímica do ponto de vista da sua actividade enzimática.

O principal objectivo deste trabalho consistiu, então, na optimização de protocolos que permitam a expressão e purificação de quatro proteínas do CFAV em *E. coli* na qual vieram a ser produzidas como proteína de fusão com “caudas” (*tags*) de hexahistidina nos seus extremos C-terminais.

## ***2. MATERIAL E MÉTODOS***

---

## 2. MATERIAL E MÉTODOS

### 2.1. Expansão do CFAV

A expansão do CFAV (gentilmente cedido por Victor Stollar, Department of Molecular Genetics, Microbiology & Immunology, UMDNJ Robert Wood Johnson Medical School, New Jersey-USA) foi feita com base na utilização de métodos convencionalmente utilizados em virologia para a replicação de vírus usando culturas celulares, e fazendo uso de uma linha C6/36 de células de *Aedes albopictus* (gentilmente cedida por Anna-Bella Failloux, Unité de Génétique Moléculaire des Bunyavirus, Institut Pasteur, Paris). Estas células foram mantidas a 28°C (na ausência de CO<sub>2</sub>) em meio Leibovitz L-15 (Lonza, EUA) suplementado com 10% de soro fetal de bovino (*fetal bovine serum* - FBS, Lonza, EUA), 200 mM de L-glutamina (Gibco BRL, EUA), 1000 U/ml de uma mistura de penicilina e streptomina (Gibco BRL, EUA), e meio de triptose fosfato (AppliChem GmbH, Alemanha) diluído 1:50. Para a replicação do CFAV, a concentração de FBS foi reduzida para 5%. Após 5 dias, procedeu-se à recolha dos sedimentos de células infectadas com CFAV, bem como do sobrenadante contendo o vírus, para posteriores utilizações.

### 2.2. Extracção de DNA total a partir de sedimentos celulares

Para a extracção de DNA total a partir de sedimentos celulares mencionados em 2.1. foi utilizado um protocolo experimental que contemplou a lise celular por acção de detergentes e a remoção de proteínas contaminantes com uma mistura de fenol e clorofórmio. Basicamente, as suspensões celulares foram recolhidas para tubos Falcon (15 ml) e centrifugadas a aproximadamente 1000g durante 15 minutos a 4°C.

Após a centrifugação, as células foram ressuspensas em 500 µl de PBS [8% (m/v) NaCl; 0,2% (m/v) KCl; 1,15% (m/v) Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O; 0,2% (m/v) KH<sub>2</sub>PO<sub>4</sub>; pH 6,5] e novamente sedimentadas por centrifugação durante 10 minutos a 2300g, e a 4°C. Em seguida, adicionaram-se 300 µl de tampão de lise (50 mM KCl, 10 mM Tris-HCl pH 8,3, 2,5 mM MgCl<sub>2</sub>, 0,1 mg/ml de gelatina, 0,45% (v/v) NP40 e 0,45% (v/v) Tween 20), tendo a inversão sucessiva dos mesmos estimulando a lise das células. Os tubos foram centrifugados a 16000g durante 2 minutos à temperatura ambiente. Após esta

centrifugação, foram adicionados 400 µl de uma mistura de fenol, clorofórmio e álcool isoamílico (nas proporções de 25:24:1) a cada tubo, os quais foram agitados de forma a promover a formação de uma emulsão e, em seguida, centrifugados nas mesmas condições descritas anteriormente. O sobrenadante foi, então, transferido para novos tubos contendo 400 µl de isopropanol, e o seu conteúdo homogeneizado por inversão suave. Em seguida, os ácidos nucleicos precipitados foram colhidos por centrifugação a 16000g durante 30 minutos à temperatura ambiente. O sedimento de ácidos nucleicos foi lavado com 250 µl de etanol 70% e centrifugado a 16000g durante 5 minutos à temperatura ambiente. Decantou-se o sobrenadante e os sedimentos foram secos sob vácuo (Concentrator 5301–Eppendorf, EUA). Por fim, o sedimento seco foi ressuspenso em 30 µl de TE (25 mM Tris-HCl pH 8,0; 10 mM EDTA) com RNase (10 µg/ml) e guardados a -20°C.

### 2.2.1. Amplificação de fragmentos de DNA por reacção em cadeia da polimerase (PCR)

Para amplificação de fragmentos de DNA por PCR foram utilizados dois sistemas: *illustra™ PuReTaq™ Ready-To-Go™ PCR Beads* (GE Healthcare, Reino Unido) e o sistema *Phusion™ High-Fidelity DNA Polymerase* (Finnzymes, Finlândia), segundo as instruções recomendadas pelos fabricantes.

O sistema *illustra™ PuReTaq™ Ready-To-Go™ PCR Beads* apresenta-se sob a forma de esferas liofilizadas (*bead*) as quais, quando reconstituída num volume final de 25 µl dão origem a uma mistura de reacção constituída em 200 µM de cada dNTP, 10 mM Tris-HCl, pH 9,0, 50 mM KCl, 1,5 mM MgCl<sub>2</sub> e 2,5 unidades de uma polimerase de DNA. A reconstituição das esferas liofilizadas foi feita com adição de 1 µl de cada um dos *primers* ITS1-F (5'-GTCGTAACAAGGTTTCCGTAGG-3')/ITS2-R (5'-GTCGGTGTTCATGTGTCCTG-3') (Freitas F. e Pinto J., comunicação pessoal) com concentração final de 10 pmol/µl e água MilliQ. O passo seguinte consistiu na adição de 2 µl de DNA extraído de sedimentos celulares, perfazendo um volume final de reacção de 25 µl. As condições da reacção de PCR utilizadas estão representadas na tabela 2.1.

**Tabela 2.1** – Condições utilizadas nas amplificações, por PCR, dos fragmentos de DNA (sedimentos celulares) da região ITS do genoma das células C6/36.

Etapa	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	94°C	2 min	1
Desnaturação	94°C	30 seg	
Hibridação	55°C	45 seg	35
Extensão	72°C	45 seg	
Extensão final	72°C	5 min	1
Conservação	4° C	∞	1

O sistema *Phusion<sup>TM</sup> High-Fidelity DNA Polymerase*, possui inúmeras vantagens, nomeadamente, a utilização de um enzima *Phusion DNA Polymerase* que é capaz de amplificar fragmentos de DNA com uma eficiência e fidelidade (anteriormente inatingíveis com uma única enzima) devido à sua capacidade de correcção de erros, a qual resulta da combinação de actividade polimerase (5'→3') e exonuclease (3'→5') de DNA. Esta enzima gera produtos de amplificação com extremidades cegas.

Resumidamente, as misturas de reacção foram preparadas para num volume final de 50 µl, contendo 10 µl de tampão de reacção *Phusion GC* concentrado 5x (suplementado com 1,5 mM MgCl<sub>2</sub>), 1 µl de 10 mM dNTP, 1 µl de cada um dos respectivos *primers* (concentração final de 10 pmol/µl), 5 µl de cDNA (ver secção 2.3.1), 0,5 µl de *Phusion DNA polymerase* (para uma concentração final de 2 U/µl) e água *MilliQ*. As condições da reacção de PCR utilizadas estão representadas na tabela 2.2.

**Tabela 2.2** – Condições utilizadas nas amplificações, por PCR, dos fragmentos de cDNA das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do invólucro, domínio helicase da proteína NS3 e NS5B).

Etapa	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	98°C	30 seg	1
Desnaturação	98°C	10 seg	
Hibridação	50°C	30 seg	35
Extensão	72°C	1:30 min	
Extensão final	72°C	10 min	1
Conservação	4° C	∞	1

A preparação das misturas de reacção e as adições de DNA às mesmas foram efectuadas em compartimentos separados. Em todas as reacções de PCR foram contemplados controlos negativos de modo a avaliar possíveis contaminações durante o processo de amplificação. As reacções de PCR decorreram num termociclador *myCycler* (Bio-Rad, EUA).

### 2.2.2. Visualização dos produtos de PCR

Os fragmentos de DNA amplificados por PCR foram visualizados após terem sido separados por electroforese em gel de agarose a 1% (em solução TAE 0,5x) na presença de 0,5 µg/ml de brometo de etídio. Os tamanhos dos diferentes fragmentos de amplificação obtidos foram estimados por comparação da sua migração electroforética com a de um marcador de massa molecular contendo moléculas lineares de DNA em cadeia dupla com tamanhos a variar de 100 pb (*GeneRuler™ 100pb Plus DNA ladder*; Fermentas, Lituânia). Para a visualização e registo em suporte digital de imagens sob luz ultravioleta (UV) dos produtos amplificados (válido para todo trabalho executado) utilizou-se o sistema *Gel Doc XR System* (Bio-Rad, EUA).

### 2.3. Extracção do RNA viral

A extracção de RNA viral a partir de sobrenadante de cultura de células C6/36 infectadas pelo CFAV, foi efectuada através da utilização do sistema comercial *INSTANT Virus RNA Kit* (AnalytikJena, Alemanha). O protocolo experimental foi realizado de acordo com as instruções recomendadas pelo fabricante. Basicamente, este

pressupôs o descongelamento de alíquotas de sobrenadante de cultura (previamente conservadas a  $-80^{\circ}\text{C}$ ) à temperatura ambiente, seguido de um passo de lise dos viriões em tubos de extracção, lavagem das colunas onde se fixa o RNA e, por fim, a sua eluição com água livre de RNAses. O volume final de cada extracção (30  $\mu\text{l}$ ) foi dividido em alíquotas de 5  $\mu\text{l}$ , as quais foram conservadas à temperatura de  $-80^{\circ}\text{C}$  até posterior utilização.

### 2.3.1. Síntese, por transcrição reversa (RT), de DNA complementar (cDNA) a partir de RNA viral

Para a síntese de cDNA a partir das moléculas de RNA viral, extraídas como descrito na secção anterior, foi utilizado o sistema comercial *RevertAid<sup>TM</sup> H Minus First Strand cDNA synthesis Kit* (Fermentas, Lituânia). O protocolo experimental foi realizado de acordo com as instruções recomendadas pelo fabricante. Genericamente, a 5  $\mu\text{l}$  de extracto de RNA foram adicionados 1  $\mu\text{l}$  de oligonucleótidos (hexâmeros) sintéticos iniciadores com sequências aleatórias, e 6  $\mu\text{l}$  de água tratada com dietil pirocarbonato (DEPC). A mistura foi incubada à  $70^{\circ}\text{C}$  durante 5 minutos (desnaturação do RNA para eliminar estruturas secundárias que as moléculas de RNA viral eventualmente formem) e, em seguida, colocada em gelo durante igual período de tempo. Seguidamente, foi preparada uma mistura de reacção contendo 4  $\mu\text{l}$  solução de tampão concentrada 5x, 1  $\mu\text{l}$  inibidor de RNases (*RiboLock<sup>TM</sup> RNase Inhibitor*, 20 U/ $\mu\text{l}$ ), 2  $\mu\text{l}$  dNTPs (concentração final de 10 mM) e 1  $\mu\text{l}$  de transcriptase reversa (*RevertAid<sup>TM</sup> H Minus M-MuLV*, 200 U/ $\mu\text{l}$ ) perfazendo um volume total de 20  $\mu\text{l}$ . As reacções de transcrição reversa foram efectuadas num termociclador *myCycler* de acordo com o programa representado na Tabela 2.3.

**Tabela 2.3** – Condições utilizadas para as reacções de transcrição reversa a partir de RNA viral.

Etapa	Temperatura ( $^{\circ}\text{C}$ )	Tempo (min)	Número de ciclos
Hibridação	25	5	1
Extensão	42	60	1
Inactivação da enzima RT	70	5	1

Os produtos resultantes da transcrição reversa (cDNA) foram conservados a -20°C até posterior utilização.

#### 2.4. Delineamento de oligonucleótidos sintéticos (*primers*) iniciadores de PCR

O desenho dos *primers* para amplificação, por PCR, de regiões da sequência nucleotídica do CFAV que codificam a proteína estrutural da cápside (C), a glicoproteína do invólucro (E), e as proteínas não estruturais NS3 (domínio com que confere à proteína actividade de helicase de RNA; a partir daqui designada NS3hel) e NS5B (polimerase de RNA dependente de RNA), foi elaborado manualmente, com base a sequência nucleotídica do CFAV, disponível na base de acesso público (acedido através do portal do NCBI, *National Center for Biotechnology Information* em [http://www.ncbi.nlm.nih.gov/nuccore/NC\\_001564.1](http://www.ncbi.nlm.nih.gov/nuccore/NC_001564.1)) com o número NC\_001564.

As sequências complementares aos *primers* que permitiram amplificar as sequências codificantes das proteínas estruturais C e E foram identificadas de forma a eliminar dos produtos amplificados as regiões que codificam secções hidrofóbicas destas proteínas, com o intuito de melhorar a sua expressão em *Escherichia coli*. De modo semelhante, o desenho dos *primers* para a amplificação de NS3hel teve por base a delimitação do domínio responsável pela sua actividade de helicase de RNA (utilizando a ferramenta CDART: *Conserved Domain Architecture Retrieval Tool*, disponível em <http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>) na sequência de NS3 do CFAV. Por fim, o par de *primers* complementar à sequência codificante de NS5B foi desenhado de forma a permitir a amplificação da totalidade da mesma.

Para além disso, na extremidade 5' de cada um dos *primers* foram adicionados nucleótidos não complementares às sequências virais (vulgarmente designados por *tag*) contendo sequências nucleotídicas que, nos produtos amplificados, corresponderiam a locais de clivagem por enzimas de restrição do tipo II. Estas sequências são reconhecidas pelos enzimas *NcoI*, *NotI* e *AseI/VspI*, os quais foram escolhidos tendo em conta a sua ausência nas sequências virais amplificadas, bem como os locais de restrição enzimática disponíveis nas sequências dos vectores de expressão utilizados. Neste caso, foram utilizados os enzimas *NcoI* e *NotI* quando o vector de clonagem/expressão foi o pET-28b (Novagen, EUA) e os enzimas *NotI* e *AseI/VspI* (compatível com o enzima



*NdeI*) quando o vector utilizado foi o pET-29b (Novagen, EUA). Toda informação acerca dos *primers* utilizados encontra-se representada na tabela 2.4.

**Tabela 2.4** – Informação relativa ao conjunto de *primers* utilizados nas amplificações, por PCR, de cDNA de CFAV efectuados ao longo deste trabalho.

Região do genoma	Nome do primer	Sequência de primer (5'---3')	Tamanho (nt)	TM <sup>1</sup> (°C)	Posição NC_001564
C	CFor <i>NcoI</i>	<b>CCGGTCGCCACC</b> ATGGAGAGGAAG GATTTAGAGGCGCG	38	68,50	114-139
	CRev <i>NotI</i>	<b>ATATATATTGCGGCCGC</b> ACGTCGGG ATAATGAGCGAATTCGC	42	74,65	410-435
E	EFor <i>NcoI</i>	<b>CCGGTCGCCACC</b> ATGGTAAAGGGG GAGTTTGTGAACCG	39	73,08	939-962
	ERev <i>NotI</i>	<b>ATATATATTGCGGCCGC</b> ACCAACCG AAAAC TTCCGACCAG	41	71,63	2094-2119
NS3hel	NS3helFor <i>AseI</i>	<b>CCGGTCGCCACC</b> ATTAATGGAGAGC TTGAACTGTGATTTCGACC	44	71,94	4992-5015
	NS3helRev <i>NotI</i>	<b>ATATATATTGCGGCCGC</b> GCGGCAG TTCATGTAGAAATCAATG	42	69,31	6203-6228
NS5B	NS5BFor <i>AseI</i>	<b>CCGGTCGCCACC</b> ATTAATGGCTCTG GAAAAGTCAACAACGATC	43	71,54	7473-7496
	NS5BRev <i>NotI</i>	<b>ATATATATTGCGGCCGC</b> CATCATTTTC CTCTTCGATATCGACTC	42	65,50	10106-10131

Todos os *primers* foram adquiridos comercialmente à empresa STAB-VIDA (Portugal).

**Nota:** <sup>1</sup> TM: temperatura de desnaturação (do inglês *temperature of melting*). A temperatura de desnaturação dos *primers* foi calculada com recurso à ferramenta disponível em [https://www.finnzymes.fi/tm\\_determination.html](https://www.finnzymes.fi/tm_determination.html), e a partir da sequência sem marcação a negrito. O início da sequência destacada a negrito corresponde a sequência de reconhecimento de enzima (s) de restrição (*tag*). O nt corresponde a “nucleótidos”.

Na tentativa de amplificação de sequências virais a partir do DNA total extraído dos sedimentos celulares utilizando *primers* específicos para o CFAV foi utilizado o sistema *illustra<sup>TM</sup> PuReTaq<sup>TM</sup>Ready-To-Go<sup>TM</sup> PCR Beads* (GE Healthcare, Reino Unido), tal como descrito na sessão 2.2.1.

Os produtos amplificados foram visualizados após terem sido separados por electroforese em gel de agarose a 1,5%, tal como descrito na sessão 2.2.2.

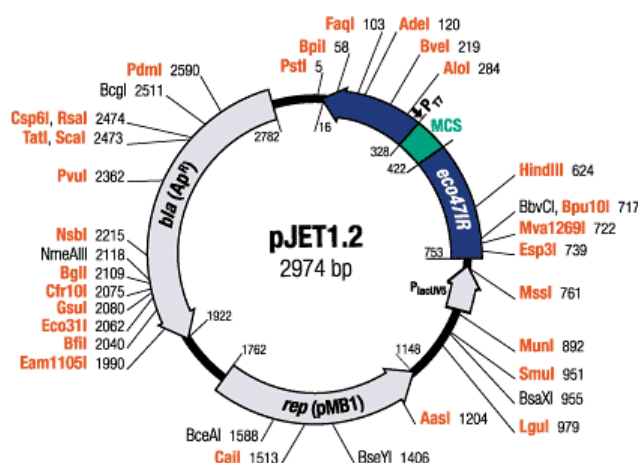
## **2.5. Amplificação de sequências virais a partir de cDNA**

A amplificação das regiões codificantes das proteínas C, E, NS3hel e NS5B foi realizada a partir de cDNA viral produzido como descrito na secção 2.3.1., utilizando os *primers* específicos apresentados na Tabela 2.4. Em cada uma das reacções de amplificação utilizou-se o sistema *Phusion<sup>TM</sup> High-Fidelity DNA Polymerase*, tal como descrito na secção 2.2.1 Os produtos amplificados foram visualizados após terem sido separados por electroforese em gel de agarose a 1,5%, tal como descrito na sessão 2.2.2.

## **2.6. Clonagem molecular de fragmentos do genoma do CFAV no vector pJET1.2/*blunt***

O vector pJET1.2/*blunt* (Fermentas, Lituânia) foi utilizado no presente estudo. Os quatro fragmentos de DNA amplificados por PCR a partir de cDNA produzido tal como descrito na secção 2.3.1, foram independentemente clonados no vector pJET1.2/*blunt*. A inserção dos genes de interesse neste vector permitiu a estabilização das sequências codificantes de proteínas do CFAV num vector plasmídico, levantando quaisquer eventuais limitações técnicas impostas pela limitação às quantidades de cDNA disponíveis para posterior clonagem nos vectores de expressão. O pJET1.2/*blunt* é obtido comercialmente sob a forma de moléculas lineares com extremidades cegas. Como principais características, o vector contém o promotor de transcrição do fago T7, um gene *bla* que codifica para a beta-lactamase (e cuja presença confere resistência à ampicilina), e duas regiões com sequências reconhecidas por múltiplas endonucleases de restrição, as quais ladeiam o local onde se podem ser ligados produtos de PCR com extremidades cegas. Possui também um gene “letal” (codificando uma endonuclease de

restrição de corte frequente) que pode ser interrompido por ligação de um inserto de DNA no local de clonagem. Como resultado, apenas as células com plasmídeos recombinante são capazes de se propagar, o que facilita o despiste de clones desejados. O mapa e características do vector pJET1.2/*blunt* encontram-se esquematizados na figura 2.1.



**Figura 2.1.** Esquema representativo do vector pJET1.2/*blunt*. (<http://www.fermentas.com/en/support/technical-reference/phage-plasmid-dna/pjet12>).

Para a ligação de DNA a este vector, foi utilizada uma razão molar inserto/vector de 3:1. As misturas de ligação foram preparadas de forma a conterem 10 µl de solução tampão de ligação concentrada 2x, 1 µl do vector (50 ng/µl), 1 µl de ligase de DNA do fago T4 e água *MilliQ* para perfazer um volume final de 20 µl. Estas misturas foram incubadas à temperatura ambiente durante 30 minutos, tendo sido posteriormente utilizadas na transformação das células competentes de *E. coli NovaBlue* (Novagen, EUA).

### 2.6.1. Preparação de células competentes de *E. coli* e sua transformação por choque térmico

A preparação de células competentes de *E. coli* foi efectuada com base numa versão ligeiramente modificada de um protocolo inicialmente descrito na bibliografia por Chung e colaboradores (Chung *et al.*, 1989). O processo foi iniciado através da obtenção de uma cultura saturada da estirpe *NovaBlue*. Primeiro, foi inoculada em 3 ml de meio líquido de crescimento de Luria-Bertani (LB; Sambrook *et al.*, 1989), a partir

de uma suspensão celular (pura e concentrada) conservada a  $-80^{\circ}\text{C}$ , seguida da sua incubação a  $37^{\circ}\text{C}$  com agitação durante a noite. Após esse período, 50  $\mu\text{l}$  da suspensão celular obtida foram diluídas em 15 ml do mesmo meio de cultura, tendo sido incubadas nas condições de crescimento descritas em cima, até que as mesmas apresentassem uma densidade óptica a variar entre 0,3-0,5 (leitura a 600 nm). Em seguida, as células bacterianas foram colhidas por centrifugação a 4500g, durante 10 minutos, a  $4^{\circ}\text{C}$ , tendo sido desprezado o sobrenadante. O sedimento celular foi directamente ressuspensionado em 500  $\mu\text{l}$  de TSS [do inglês, *transformation and storage solution*; Lb 1x, 10% (v/v) PEG 6000, 5% (v/v) DMSO, 50 mM  $\text{MgSO}_4$ , pH 6,5] mantido a  $4^{\circ}\text{C}$ .

Para a transformação das células competentes, adicionaram-se 10  $\mu\text{l}$  de cada mistura de ligação a 200  $\mu\text{l}$  das células ressuspensionadas em meio TSS em tubos de 1,5 ml frios. As suspensões obtidas foram mantidas em gelo durante 1 hora. Seguidamente estas foram submetidas a um choque térmico a  $42^{\circ}\text{C}$  durante 90 segundos, e imediatamente recolocados 10 minutos em gelo. Em seguida, foi adicionado a cada tubo um volume de 900  $\mu\text{l}$  de meio SOC (*Super Optimal broth for Catabolite repression*; Sambrook *et al.*, 1989), e mantidas 1 hora a  $37^{\circ}\text{C}$ , com agitação suave. Findo este período de tempo, distribuíram-se 200  $\mu\text{l}$  da suspensão bacteriana à superfície de caixas de meio LB sólido suplementado com antibióticos [ampicilina (100  $\mu\text{g/ml}$ )/tetraciclina (12,5  $\mu\text{g/ml}$ ) ou canamicina (30  $\mu\text{g/ml}$ )/ cloranfenicol (25  $\mu\text{g/ml}$ ), dependendo da estirpe e do vector utilizado]. As caixas foram incubadas a  $37^{\circ}\text{C}$  numa estufa, durante a noite (entre 16 a 18 horas) de forma a permitir o crescimento de colónias bacterianas.

### 2.6.2. Extracção do DNA plasmídico por lise alcalina

A extracção do DNA plasmídico foi executada através da lise alcalina de suspensões bacterianas. As colónias identificadas como possíveis clones recombinantes, obtidos pela inserção de DNA alvo nos vectores utilizados, foram repicadas, com um palito estéril, para tubos com 3 ml de meio LB líquido suplementado com antibióticos [ampicilina (100  $\mu\text{g/ml}$ )/tetraciclina (12,5  $\mu\text{g/ml}$ ) ou canamicina (30  $\mu\text{g/ml}$ )/ cloranfenicol (5  $\mu\text{g/ml}$ ), dependendo do vector utilizado]. Estes foram mantidos durante a noite a  $37^{\circ}\text{C}$ , com agitação, até à obtenção de uma cultura bacteriana saturada. Posteriormente, centrifugou-se 2 ml da mesma a 16000g durante 2 minutos, tendo o

sedimento celular sido recuperado, após decantação do sobrenadante. As células foram, posteriormente, ressuspensas em 200 µl de TEG (25 mM Tris-HCl pH 8,0, 1 mM EDTA, 1% de glucose). Em seguida, adicionaram-se 250 µl de solução de lise (0,2 M NaOH, 1,5% SDS) e procedeu-se à agitação suave dos tubos, por inversão, para estimular a lise das células bacterianas. Em seguida adicionaram-se 250 µl de acetato de potássio (3 M, pH 5,4) e procedeu-se, novamente, à sua agitação por inversões sucessivas. Os tubos contendo o lisado celular foram centrifugados a 16000g durante 15 minutos à temperatura ambiente. O sobrenadante foi transferido para novos tubos contendo igual volume de isopropanol (aproximadamente 700 µl), e o seu conteúdo homogeneizado por inversão suave. Em seguida, os ácidos nucleicos precipitados foram colhidos por centrifugação nas mesmas condições descritas anteriormente. O sedimento de ácidos nucleicos foi lavado com 250 µl de etanol 70% e centrifugado a 16000g durante 5 minutos. Decantou-se o sobrenadante e os sedimentos foram secos sob vácuo. Por fim, o sedimento seco foi ressuspensado em 30 µl de TE (1 M Tris-HCl, pH 8,0, 0,5 M EDTA) com 100 µg/ml de RNase e incubado a 37°C, em banho-maria, durante 30 minutos, para permitir uma eficiente degradação do RNA presente.

Para a identificação das moléculas de DNA plasmídico potencialmente recombinantes, analisou-se a migração, em gel de agarose a 0,7% (tal como descrito em 2.2.2.) das moléculas extraídas sem que estas tenham sido digeridas com qualquer enzima de restrição. Foi observada a diferença no padrão de migração destas moléculas comparativamente à de um plasmídeo de referência (controlo negativo) com aproximadamente 3 kb, tendo sido seleccionadas as moléculas de plasmídeo com migração retardada para posterior análise por digestão enzimática.

### **2.6.3. Digestão enzimática de DNA plasmídico**

A identidade dos clones de DNA recombinante pretendidos foi confirmada pela sua clivagem enzimática com endonucleases de restrição. Para o efeito, foram preparadas misturas de reacção com extractos contendo DNA plasmídico com volumes a variar entre 1,5 a 6 µl (dependendo da quantidade obtida no processo de extracção), 0,3 µl de enzima [*Not*I e *Nco*I (C/E); *Not*I e *Ase*I/*Vsp*I (NS3hel/NS5B)], 2 µl de solução tampão concentrada 10x, e água *MilliQ* para um volume final de 20 µl. As reacções de

digestão foram incubadas a 37°C em banho-maria durante, aproximadamente, 2 horas. Uma vez terminada esta incubação, os produtos resultantes foram analisados por electroforese em gel de agarose 1%, tal com descrito em 2.2.2.

#### 2.6.4. Análise de moléculas recombinantes de DNA plasmídico através da amplificação, por PCR, dos fragmentos clonados

Quando conveniente, a confirmação da natureza recombinante das moléculas de DNA plasmídico foi efectuada através da amplificação, por PCR, dos insertos de DNA nelas contidos. Para tal, foi utilizado-se o sistema *Phusion<sup>TM</sup> High-Fidelity DNA Polymerase*, tal como descrito na secção 2.4, com algumas alterações, nomeadamente, na preparação das amostras de DNA plasmídico em que se fez uma diluição: C (1:10), E (1:50) e NS5B (1:5). As condições de amplificação utilizadas encontram-se descritas na Tabela 2.5.

**Tabela 2.5.** Condições utilizadas nas amplificações, por PCR, dos fragmentos de DNA das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do invólucro, NS3hel e NS5B).

Etapa	Temperatura (°C)	Tempo	Número de ciclos
<b>Desnaturação inicial</b>	95	2 min	1
<b>Desnaturação</b>	95	10 seg	
<b>Hibridação</b>	50	45 seg	40
<b>Extensão</b>	72	1:30 min	
<b>Extensão final</b>	72	5 min	1
<b>Conservação</b>	4	∞	1

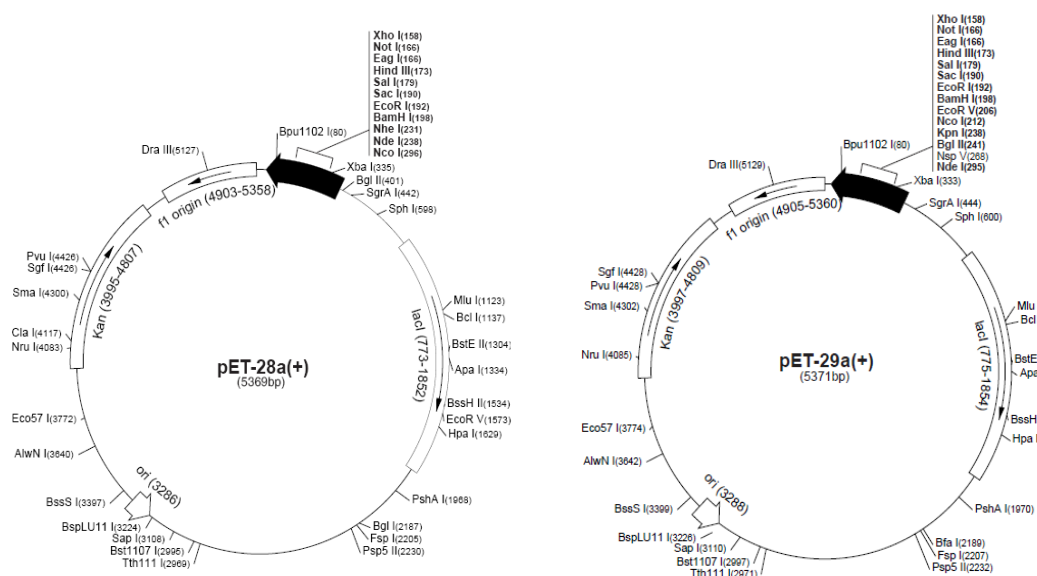
Os produtos amplificados foram visualizados após terem sido separados por electroforese em gel de agarose a 1,5%, tal como descrito na sessão 2.2.2.

### 2.6.5. Purificação dos produtos de PCR

Os produtos obtidos por PCR foram purificados de forma a puderem ser, em seguida, utilizados na reacção de ligação aos vectores de expressão. Para tal, utilizou-se um sistema comercial *DNA Clean & Concentrator*<sup>TM</sup>-5 (Zymo Research, EUA). O protocolo experimental foi realizado de acordo com as instruções recomendadas pelo fabricante. Basicamente, o protocolo de purificação emprega um sistema de um único tampão que permite a adsorção eficaz de DNA pela matriz, lavagem das colunas e, por fim, a eluição dos ácidos nucleicos purificados. A eluição do DNA foi feita com 20 µl de tampão Tris-HCl (pH 8,0) diluído a 10 mM. Uma vez terminada a purificação, os produtos resultantes foram analisados por electroforese em gel de agarose 1%, tal com descrito em 2.2.

### 2.7. Preparação dos vectores de expressão pET-28b e pET-29b

Os vectores de expressão utilizados neste trabalho foram o pET-28b e o pET-29b (Novagen, EUA). Como principais características, estes vectores contêm um promotor reconhecido pela polimerase de RNA do fago T7 (o qual não é reconhecido pela polimerase de RNA de *E. coli*), uma sequência do operador *lac* (*lacI*), o local de clonagem múltipla e um gene que confere resistência à canamicina (*kan*). Os mapas de restrição, as características e as sequências com os locais de ligação dos insertos, dos vectores pET-28b e pET-29b encontram-se esquematizados na figura 2.2.





**Figura 2.2.** Esquemas representativos dos mapas dos vetores pET-28a e pET-29a (Novagen, EUA). Os mapas mostram a origem de replicação (*ori*), um local de clonagem múltipla (seta a negrito) e um gene que confere resistência a canamicina (*kan*) (<http://www.merck-chemicals.com/life-science-research/vector-table-novagen-pet-vector-table/cHdSb.s1O77QAAAEhPqsLd cab>).

### 2.7.1. Digestão enzimática dos produtos de PCR purificados e dos vetores de expressão pET-28b e pET-29b

Para a construção dos plasmídeos de expressão contendo os insertos dos genes codificantes das proteínas C, E, NS3hel e NS5B, obtidos conforme descrito anteriormente, foram realizadas digestões enzimáticas. Os fragmentos codificantes das proteínas C e E e o vector pET-28b foram duplamente digeridos com os enzimas de restrição *Nco*I e *Not*I. Quanto aos fragmentos codificantes das proteínas NS3hel e NS5B, estes foram duplamente digeridos com os enzimas de restrição *Ase*I/*Vsp*I e *Not*I e o vector pET-29b com os enzimas de restrição *Nde*I (compatível com *Ase*I/*Vsp*I) e *Not*I.

As reacções de digestão foram efectuadas num volume final a variar entre 30 a 70 µl (dependendo da quantidade de DNA obtida no processo de purificação), 3 a 7 µl de tampão de reacção concentrado 10x e 0,75 a 2 µl de enzima de restrição (dependendo da abundância do fragmento a digerir).



As segundas reacções de digestão enzimática foram antecedidas de uma diálise das amostras de DNA previamente clivados por uma primeira enzima (insertos e vectores), contornando as incompatibilidades impostas às digestões subsequentes pela composição dos tampões de reacções de restrição utilizados inicialmente. Basicamente, numa placa de *Petri* com água *MilliQ* colocaram-se discos de nitrocelulose de 0,025 µm (Millipore, Irlanda) a flutuar e no centro de cada um deles, foram aplicadas as amostras a dialisar, tendo esta decorrido durante duas horas.

Todas as reacções de digestão enzimática foram incubadas a 37°C em banho-maria durante 2 horas. Seguidamente as amostras foram aplicadas num gel de agarose 1%, tal com descrito na secção 2.2.2. Todas as enzimas de restrição utilizadas neste trabalho são comercializadas pela Fermentas (Lituânia).

## **2.8. Clonagem molecular de fragmentos do genoma do CFAV nos vectores de expressão pET-28b e pET-29b**

Antes de se realizar as reacções de ligação dos insertos de DNA aos vectores de expressão, ambas as moléculas foram co-precipitadas (para concentração dos ácidos nucleicos) depois da inactivação, por tratamento pelo calor, dos enzimas previamente utilizados.

A inactivação das misturas de digestão enzimática decorreu a 80°C num banho seco (*Test tube Heater SHT 10* - Stuart, Reino Unido) durante 30 minutos tendo sido, posteriormente, adicionados em cada tubo água *MilliQ* (para prefazer um volume final de 100 µl), 1 µl de glicogénio (20 mg/ml; Fermentas, Lituânia), 10 µl de acetato de sódio 3 M, pH 5,4, e 300 µl de etanol 96%. De imediato, as misturas foram agitadas vigorosamente durante alguns segundos e incubadas a -80°C durante 30 minutos. Terminada a incubação, seguiu-se a centrifugação a 16000g durante 30 minutos à temperatura ambiente. Os sedimentos foram lavados com 300 µl de etanol 70% e centrifugados a 16000g durante 10 minutos à temperatura ambiente. Decantou-se o sobrenadante e os sedimentos de ácidos nucleicos foram secos sob vácuo.

A reacção de ligação foi iniciada ressuspendendo os sedimentos obtidos com 17 µl de água *MilliQ*, 2 µl de tampão adequado, concentrado 10x, e 1 µl de ligase de DNA

do fago T4, concentrada a 5 U/ $\mu$ l (Fermentas, Lituânia). As misturas foram incubadas a 16°C durante a noite e, posteriormente, guardadas a -20°C.

A preparação de células competentes de *E. coli* e sua transformação foi efectuada, tal como descrita na secção 2.6.1. A extracção do DNA plasmídico por lise alcalina foi efectuada, tal como descrita na secção 2.6.2. Para a identificação das moléculas de DNA plasmídico potencialmente recombinantes, analisou-se a migração, em gel de agarose a 1,5%, tal como descrito em 2.2.2. [mas com a utilização de outro marcador de massa molecular contendo de moléculas lineares de DNA em cadeia dupla  $\lambda$ DNA/Eco91I (BstEII); Fermentas, Lituânia].

Para confirmar a presença do inserto no plasmídeo e a ligação correcta do inserto, os clones recombinantes foram digeridos com enzimas de restrição [*AseI/VspI* (pET-28b::C e pET-29b::NS3hel/NS5B), *XhoI* (pET-28b::E) e *NcoI* (pET-29b::NS3hel)], tal como descrito na secção 2.6.3. Os produtos resultantes foram analisados por electroforese em gel de agarose 1,5%, tal com descrito em 2.2.2.

### 2.8.1. Purificação de DNA plasmídico

Os produtos obtidos pela extracção do DNA plasmídico por lise alcalina foram purificados utilizando um sistema comercial DNA *Clean & Concentrator*<sup>TM</sup>-5. O protocolo experimental foi realizado de acordo com as instruções recomendadas pelo fabricante e tal com descrito na secção 2.6.5. Uma vez terminada a purificação, os produtos resultantes foram analisados por electroforese em gel de agarose 1,5%, tal com descrito em 2.2.2.

### 2.8.2. Sequenciação de DNA

As moléculas de DNA purificadas foram sequenciadas pela empresa STAB VIDA (Costa de Caparica, Portugal), por recurso à técnica de terminação da polimerização em cadeia com didesoxi-ribonucleótidos, descrita inicialmente por Sanger (Sanger *et al.*, 1977). Foi efectuada a sequenciação parcial das moléculas de DNA clonados (região próxima dos locais onde ocorreu a ligação de insertos aos plasmídeos

utilizados como vectores), e em ambos os sentidos, utilizando *primers* universais, complementares as sequências dos vectores.

## 2.9. Expressão de proteínas recombinantes

A escolha das estirpes de expressão depende das características da proteína e do uso que se lhe destina. *E. coli* tem sido um dos sistemas bacterianos de primeira escolha, por se tratar de um microrganismo da qual se conhece o genoma, ser fácil de transformar, não ser patogénico (pelo menos não o são as estirpes comumente utilizadas em experiências de clonagem e expressão proteica) e com produtividade elevada em termos de biomassa.

Um dos passos para a expressão de proteínas recombinantes em *E. coli* foi escolher e preparar as estirpes de expressão para a sua transformação. E neste trabalho, as estirpes de *E. coli* utilizadas foram a BL21(DE3)pLysS [*F-ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm* (DE3)pLysS (Cam<sup>R</sup>)] e Rosetta(DE3)pLysS [*F-ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm* (DE3)pLysSRARE2 (Cam<sup>R</sup>)]. Ambas possuem um gene que codifica uma polimerase de RNA do fago T7 (sob o controlo transcricional do promotor *lacUV5*) integrada no genoma de uma variante não indutível do fago lambda (DE3), o qual pode ser induzido com isopropil-β-D-tiogalactopiranosídeo (IPTG). Estas estirpes apresentam, ainda, deficiência nas proteases OmpT e Lon, reduzindo a degradação das proteínas recombinantes durante a sua purificação. A presença do plasmídeo pLysS que codifica um gene que lhes confere resistência ao cloranfenicol e um gene que codifica a lisozima do fago T7 que é inibidora natural da respectiva polimerase de RNA, permite manter o nível de expressão proteica basal a um nível reduzido. A lisozima do fago T7 reduz a expressão basal dos genes alvo e, portanto, estabiliza os plasmídeos que expressam proteínas tóxicas para a *E. coli*. Para além disso, a estirpe Rosetta (DE3)pLysS tem como característica muito particular transportar genes que codificam tRNA para códons raramente utilizados em *E. coli* (AGG/Arg, AGA/Arg, AUA/Ile, CUA/Leu, CCC/Pro, e GGA/Gly). Estes genes são veiculados numa variante de plasmídeo pLysS (designado pLysSRARE), permitindo a expressão de proteínas em cuja sequência codificante estes estejam presentes.

A preparação de células competentes de *E. coli* foi efectuada, tal como descrita na secção 2.6.1., e os clones bacteriano contendo plasmídeos recombinantes foram re-inoculados e preparados para posterior conservação a -80°C, tal como descrito na secção 2.8.1.

### **2.9.1. Indução da expressão de proteínas recombinantes e avaliação da sua toxicidade através da monitorização do crescimento bacteriano em meio líquido**

Para esta etapa do trabalho, foram utilizados os protocolos, com ligeiras modificações, descritos no *pET System Manual 11<sup>th</sup> Edition* (Novagen, EUA).

A partir de uma cultura bacteriana conservada a -80°C, foi efectuado um riscado em meio LB sólido suplementado com canamicina (30 µg/ml) e cloranfenicol (25 µg/ml), pretendendo-se a obtenção de colónias isoladas. Em seguida, preparou-se um pré-inóculo em meio líquido, repicando com um palito estéril uma colónia para tubos com meio LB líquido suplementado com canamicina (30 µg/ml) e cloranfenicol (5 µg/ml). Estes foram mantidos durante a noite a 37°C, com agitação, até à obtenção de uma cultura bacteriana saturada. Seguidamente, esta cultura bacteriana foi diluída 1:100 em 20 ml de meio LB líquido suplementado com canamicina (30 µg/ml) e cloranfenicol (5 µg/ml), e mantida a 37°C com agitação, promovendo o crescimento bacteriano até ser atingida uma densidade óptica (600 nm) de aproximadamente 0,5. Neste ponto, a expressão proteica foi induzida com adição de IPTG para uma concentração final de 1 mM e as culturas foram incubadas a 30 e 37°C com agitação durante 4 horas. Ao longo deste período o crescimento da cultura foi seguido através do registo da D.O.<sub>600nm</sub> das diferentes culturas. De 30 a 30 minutos, e em paralelo com as medidas de densidade óptica, foram recolhidas alíquotas de 1 ml das diferentes culturas bacterianas, em condições assépticas. As células nelas contidas foram colhidas por centrifugação durante 1 minuto a 16000g à temperatura ambiente. Decantou-se o sobrenadante e os sedimentos celulares foram guardados a -80°C para, posteriormente, serem utilizados para avaliar o nível de expressão das proteínas.

### **2.9.2. Efeito da temperatura e concentração de indutor na expressão proteica**

Existem vários parâmetros que influenciam a expressão de proteínas recombinantes usando células procariontes e que incluem, entre outros, a estirpe bacteriana utilizada, a sua temperatura de crescimento, a concentração do agente indutor, e a fase de crescimento na qual deve ser efectuada a indução da expressão proteica.

Neste trabalho foram avaliados dois parâmetros: a temperatura de crescimento e a concentração do indutor (IPTG), que podem afectar, de forma significativa, a qualidade conformacional, o rendimento de produção e a solubilidade de proteínas recombinantes produzidas.

Foram testadas três temperaturas diferentes, 15, 25 e 37°C, e para o efeito, a indução foi realizada, tal com o descrito na secção 2.8.3., sendo que as culturas bacterianas foram induzidas com 1 mM de IPTG e incubadas durante a noite, com agitação.

Ao mesmo tempo, também foram testadas duas concentrações de indutor (IPTG), 0,4 e 1 mM. Procedeu-se a indução, tal com descrito anteriormente, excepto a temperatura de incubação que foi de 37°C (durante 4 horas e com agitação). A diferentes momentos após a adição de indutor, procedeu-se a recolha de alíquotas de 1 ml, em condições assépticas, das quais se recolheram as células bacterianas por centrifugação durante 1 minuto a 16000g à temperatura ambiente. Decantou-se o sobrenadante e os sedimentos celulares foram guardados a -80°C para, posteriormente, serem analisados por SDS-PAGE.

### **2.9.3. Preparação de extractos de proteínas solúveis totais e sua análise por SDS-PAGE**

Para a verificação da solubilidade dos extractos proteicos foi utilizado um volume de 20 ml de cultura no qual a expressão de proteínas recombinantes foi promovida por adição de 1 mM de IPTG, seguida da incubação das culturas em questão durante 4 horas a 37°C. Todo o volume submetido a uma centrifugação a 4500g durante 10 minutos a 4°C para recolha das células bacterianas. Decantou-se o sobrenadante e o sedimento foi guardado a -80°C, até serem utilizados.

O sedimento congelado foi ressuspensionado em tampão de lise frio concentrado 10x (20 mM Tris-HCl, 0,5 M NaCl, pH 7,5) numa razão de 4 ml de tampão de lise por cada 40 ml de cultura. Em seguida, as suspensões foram lisadas por sonicação num sonicador *Ultrasonic processor UP200S* (Hielscher, Alemanha) nas seguintes condições: 60% de amplitude, 0,5 ciclos e 1-10 minutos (dependendo da viscosidade da suspensão). Os lisados obtidos foram centrifugados a 14000g durante 10 minutos a 4°C, de modo, a obtermos duas fracções: solúvel (sobrenadante) e insolúvel (*pellet*).

Em seguida, procedeu-se a lavagem da fracção insolúvel com 375 µl de tampão frio concentrado 10x (20 mM Tris-HCl, 0,5 M NaCl, pH 7,5) seguida de centrifugação a 10000g durante 5 minutos a 4°C. Repetiu-se a lavagem e, no final, o sedimento foi ressuspensionado em 750 µl de 1% de SDS. Fez-se a recolha de alíquotas de 100 µl de cada uma das fracções (solúvel e insolúvel) e o restante volume foi guardado a -80°C.

Para análise de proteínas presente nos extractos proteicos obtidos anteriormente procedeu-se à ressuspensão dos sedimentos e das alíquotas congeladas em 100 µl de tampão de amostra [0,5 M Tris-HCl pH 6,8, 20% (v/v) de glicerol, 10% (m/v) SDS, 0,5% (m/v) de azul de bromofenol, 5% (m/v) de β-mercaptoetanol]. As suspensões foram levadas a ferver durante 10 minutos a 100°C e, posteriormente, aplicadas num gel de electroforese de poliacrilamida (concentração a variar entre 9 e 10%) usando um sistema de tampões contendo tricina e SDS (Haider *et al.*, 2010). Após terminada a electroforese o gel de poliacrilamida foi corado com uma solução comercial de Azul de Coomassie (*Bio-Safe™ Coomassie Blue G250 Stain*; BioRad, EUA) e incubado com agitação orbital suave, à temperatura ambiente até as bandas proteicas serem claramente visíveis. As massas moleculares das diferentes proteínas foram estimadas por comparação com a migração relativa de um marcador de proteínas com massa moleculares conhecidas com bandas de 170, 118, 75, 49, 38, 25, 18, 12 e 9 kDa (*ProSieve® Color Protein Markers*; Lonza, EUA).

## 2.10. Purificação de proteínas recombinantes com “caudas” de hexahistidina

A purificação das proteínas C, E, NS3hel e NS5B expressas em fusão com uma “cauda” de 6 resíduos de histidina foi realizada por cromatografia de afinidade. Esta foi levada a cabo através da passagem de extractos proteicos por uma coluna empacotada

com resina contendo íons níquel imobilizados. Neste trabalho foram utilizados dois sistemas comerciais: *His SpinTrap* (GE Healthcare, Reino Unido) e *Ni-NTA Spin Kit Handbook* (Qiagen, Alemanha), dependendo da disponibilidade dos mesmos no laboratório. O processo foi iniciado com a indução, tal com o descrito na secção 2.8.3.

### 2.10.1. Purificação de proteínas em condições desnaturantes

Para a purificação, em condições desnaturantes, de proteínas recombinantes com “caudas” de hexahistidina, foi utilizado o protocolo do sistema comercial *His SpinTrap* (GE Healthcare, Reino Unido), segundo as instruções do fabricante, mas com algumas modificações. Basicamente, o protocolo pressupõe a ressuspensão do sedimento celular congelado (tal como descrito na secção 2.9.3.) em tampão de lise, seguido de um passo de lise mecânica, clarificação do lisado, lavagem da coluna onde se fixa a proteína e, por fim, a eluição da proteína.

Nos tampões de ligação, lavagem e eluição foram preparados com 20 mM Tris-HCl, 500 mM de NaCl e adicionados de 1% de detergente Triton<sup>TM</sup>X-100 e 10% de glicerol. Os sedimentos celulares (correspondente a 20 ml de cultura induzida) foram ressuspensos utilizando 1,4 ml de tampão de fixação (*binding-buffer*) contendo 8 mM de ureia e 56 µl de inibidor de proteases [1 mM *cOmplete - protease inhibitor cocktail* (Roche, Alemanha)]. Em seguida, as suspensões foram lisadas por sonicação nas seguintes condições: 70% de amplitude, 0,7 ciclos e 3 a 10 minutos (dependendo da viscosidade da suspensão). Os lisados foram centrifugados a 14000g durante 10 minutos a 4°C. Os clarificados obtidos foram aplicados à coluna e, finalmente eluídos, com concentrações de imidazol a variar entre 100, 200 e 500 mM. Os extractos purificados foram guardados a -80°C para, posterior quantificação.

### 2.10.2. Purificação de proteínas em condições nativas

Para a purificação, em condições nativas, de proteínas recombinantes com caudas de hexahistidina, foi utilizado o protocolo do sistema comercial *Ni-NTA Spin Kit Handbook* (QIAGEN, Alemanha), segundo as instruções do fabricante, mas com algumas modificações. Os tampões de lise, lavagem e eluição foram preparados com 50

mM  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ , 500 mM de NaCl e adicionados de 1% de detergente Triton<sup>TM</sup>X-100 e 10% de glicerol. Os sedimentos celulares (correspondente a 20 ml de cultura induzida) foram ressuspensos com 1,4 ml de tampão de lise e 56  $\mu\text{l}$  de inibidor de proteases [1 mM *cOmplete - protease inhibitor cocktail* (Roche, Alemanha)], e incubados no gelo durante 30 minutos. Em seguida, as suspensões foram lisadas por sonicação nas mesmas condições descritas anteriormente. Os lisados foram centrifugados a 12000g durante 25 minutos a 4°C, e os clarificados obtidos foram aplicados à coluna e, finalmente eluídos com uma solução contendo 500 mM de imidazol. As amostras purificadas foram guardadas a -80°C para, posterior quantificação.

## 2.11. Detecção de proteínas por *Western Blot*

Para a identificação das proteínas recombinantes com “caudas” de hexahistidina foi utilizado o método *Western Blot*. Basicamente, esta técnica pressupõe a separação de proteínas por electroforese, a sua transferência para uma membrana, o bloqueio da mesma com uma solução proteica, a detecção de proteínas imobilizadas com anticorpos (primário e secundário) e, finalmente, a detecção de complexos antigénio/anticorpo.

### 2.11.1. Transferência de proteínas para a membrana de nitrocelulose

A transferência de proteínas separadas por SDS-PAGE de um gel de acrilamida para uma membrana de *Nitrocellulose Hybond<sup>TM</sup>ECL<sup>TM</sup>* (Amersham Pharmacia Biotech, Reino Unido) foi efectuada mediante a utilização do sistema *Trans-Blot SD Semi-Dry Transfer Cell* (Bio-Rad, EUA), segundo as instruções do fabricante. Após a transferência (sistema “semi-seco”), a presença de proteínas na membrana de nitrocelulose foi verificada usando uma solução de *Rouge Ponceau* (CHROMA-Gesellschaft Schmid GMBH & CO, Alemanha). Esta solução de coloração foi removida através da lavagem da membrana com uma solução 0,1 M de NaOH e água destilada.



### 2.11.2. Bloqueio e incubação com anticorpos

Para o bloqueio da membrana de nitrocelulose foi utilizada uma solução de TBST (10 mM Tris-HCl; 150 mM NaCl; 0,05% Tween 20; pH 8,0) contendo 10% leite magro (*Blotting Grade Blocker NonFat Dry Milk*; Bio-Rad, EUA). A membrana foi incubada durante a noite a 4°C. Após a incubação, a membrana foi lavada 3x durante 5 minutos com solução de TBST, para retirar o excesso da solução de bloqueio. Em seguida, a membrana foi colocada a incubar com 2 µl de anticorpo primário (*monoclonal anti-polyhistidin antibody produced in mouse*; Sigma-Aldrich, EUA) em TBST contendo 2% leite magro, diluído 1:10000, com agitação suave durante 2 horas à temperatura ambiente. Decantou-se a solução contendo o anticorpo primário e a membrana foi lavada 3x 5 minutos com TBST. A membrana foi colocada, novamente, a incubar com 2 µl de anticorpo secundário (*anti-mouse IgG peroxidase*; Sigma-Aldrich, EUA) em TBST contendo 2% leite magro, diluído 1:5000, com agitação suave durante 1 hora à temperatura ambiente. Decantou-se a solução contendo o anticorpo secundário e a membrana foi lavada 4x 5 minutos com TBST.

Para a detecção dos complexos antigénio-anticorpo foi utilizado o sistema comercial *SIGMAFAST<sup>TM</sup> DAB With Metal Enhancer* (Sigma-Aldrich, EUA), segundo as instruções do fabricante. A comparação entre a membrana e o gel permite, assim, a identificação das proteínas relevantes e, eventualmente, o isolamento e caracterização.

### 2.12. Quantificação das proteínas

A quantificação de proteínas nos extractos obtidos após cromatografia de afinidade foi efectuada utilizando um sistema comercial *Protein Assay* (Bio-Rad, EUA), segundo as instruções do fabricante. O ensaio foi realizado em duplicado e através da leitura da densidade óptica das amostras a 595 nm ( $D.O._{595 \text{ nm}}$ ) num espectrofotómetro UV/Vis *Smart Spec<sup>TM</sup> 3000* (Bio-Rad, EUA). O valor de absorvância obtido, em cada amostra, foi utilizado para calcular a concentração de proteína purificada (mg/ml) através de uma recta padrão de densidade óptica *versus* concentração de proteína. Esta recta padrão foi construída utilizando concentrações conhecidas de gama-globulina sérica de bovino.

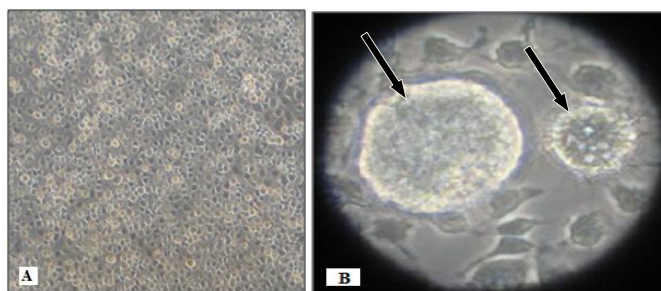
### ***3. RESULTADOS***

---

### 3. RESULTADOS

#### 3.1. Expansão do CFAV em cultura de células C6/36

O vírus CFA (CFAV) foi expandido (para obtenção de uma suspensão viral com um título elevado) utilizando a linha celular C6/36 (derivada de *Aedes albopictus*), conforme procedimento descrito na secção 2.1. do capítulo anterior. Após uma passagem (com redução da concentração de FBS no meio de cultura de 10% para 5%) e cinco dias de incubação, foi possível observar o efeito citopático que melhor caracteriza a replicação do CFAV nestas células, e do qual deriva a designação deste vírus. De facto, e tal como se pode verificar na Figura 3.1.B, as células C6/36 infectadas com o CFAV fundem-se (muito provavelmente entre si e com células não infectadas), dando origem à formação de sincícios de dimensões variáveis.



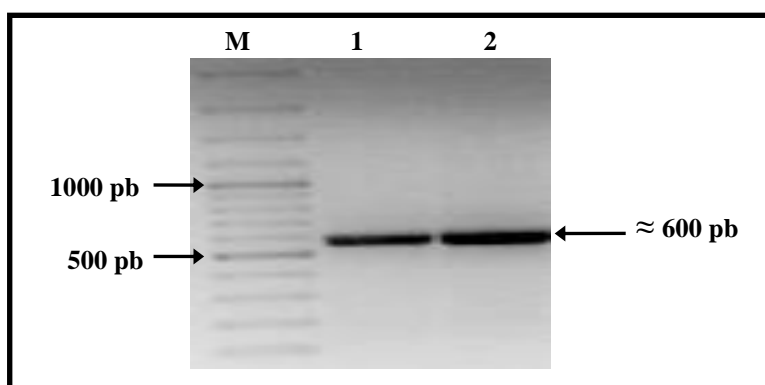
**Figura 3.1.** Observação ao microscópico óptico de células C6/36 não infectadas (A; x100) ou infectadas com o CFAV (B; x400). As setas indicam os sincícios de diferentes dimensões.

Findo o período de incubação acima mencionado, procedeu-se à recolha dos sedimentos celulares e dos respectivos sobrenadante de culturas de células C6/36 infectadas com o CFAV, os quais serviram de material de partida para a extracção de DNA genómico total (sedimentos celulares) e de RNA viral (partindo dos sobrenadantes de cultura). Com este material, pretendemos, inicialmente, verificar se os *primers* desenhados (secção 2.4. do capítulo anterior) permitiam a amplificação de sequências virais a partir do cDNA que irá ser preparado a partir de RNA viral, e se esses mesmos *primers* permitiam, ou não a amplificação de sequências de um vírus defectivo, designado *Cell Silence Agent* (CSA), integrado sob a forma de múltiplos fragmentos de DNA nas células C6/36 (Crochu *et al.*, 2004). Antes, porém, de verificarmos estes dois passos, optámos por nos assegurarmos da qualidade da matriz a ser utilizada. Para tal, procedeu-se a utilização de *primers* genéricos para a região ITS dos operões de rRNA

dos mosquitos e constatar a possibilidade de amplificação de DNA a partir dos sedimentos celulares obtidos.

### 3.2. Amplificação de fragmentos de DNA a partir dos sedimentos de células C6/36, por reacção em cadeia da polimerase (PCR)

Os sedimentos de células C6/36 infectadas, ou não, com o CFAV foram recolhidos e utilizados para extracção de DNA total. Uma porção do DNA total extraído foi, posteriormente, amplificada por PCR (região ITS1 do rRNA; ver secção 2.2.1.) com o objectivo de verificarmos a sua integridade, tal como acima descrito. O resultado da separação por electroforese em gel de agarose 1% comprovou a amplificação de um fragmento de DNA com tamanho de aproximadamente 600pb (Figura 3.2).

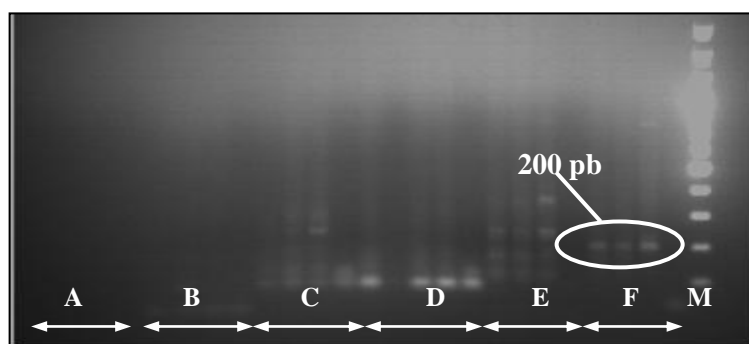


**Figura 3.2.** Separação por electroforese em gel de agarose dos produtos de PCR obtidos usando *primers* complementares à região ITS1 do operão de rRNA e DNA extraído de células C6/36 infectadas, ou não, com o CFAV. (M) marcador de massa molecular, representado à esquerda (*GeneRuler<sup>TM</sup> 100pb Plus DNA ladder*), e fragmento, de aproximadamente 600pb, correspondente ao produto de amplificação obtido usando extractos de DNA de: (1) células C6/36 não infectadas; (2) células C6/36 infectadas com o CFAV.

Subsequentemente, procedeu-se à tentativa de amplificação de sequências virais a partir do DNA total extraído dos sedimentos celulares utilizando *primers* específicos para o CFAV. Estes *primers* amplificam secções do genoma do CFAV que codificam as diferentes proteínas que nos propusemos produzir enquanto moléculas recombinantes usando *E. coli* como hospedeira de expressão. Foram elas a proteína estrutural da cápside (C), a glicoproteína estrutural de superfície (E) (ambas desprovidas de domínios

hidrofóbicos de inserção em membranas), e as proteínas não estruturais NS3hel (domínio com que confere à proteína actividade de helicase de RNA) e NS5B (polimerase de RNA dependente de RNA).

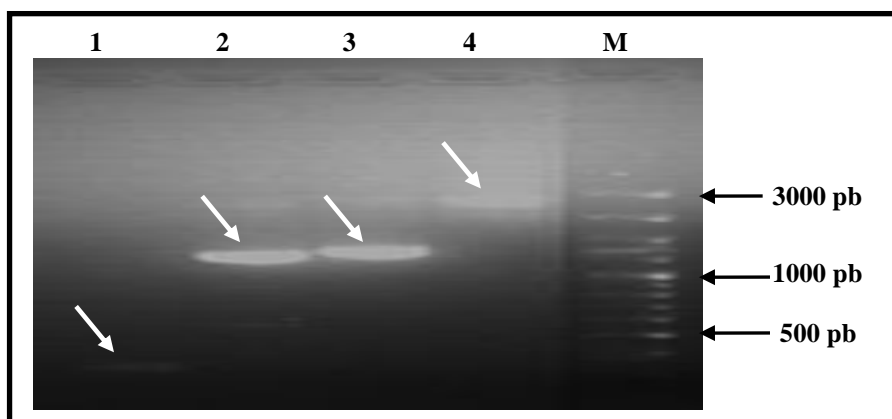
A análise dos produtos obtidos como resultado do processo de amplificação revelou (em gel de agarose) a ausência de DNA amplificado especificamente, sugerindo que, durante a sua replicação em células C6/36, o CFAV não se integrou no genoma das mesmas. Por outro lado, os *primers* desenhados parecem ser específicos para o CFAV (como esperado) não permitindo a amplificação de fragmentos do CSA. Assim, a ausência dos produtos de amplificação esperados usando extractos de DNA de células C6/36 infectadas com CFAV traduz ainda a especificidade (ainda que parcial no caso dos *primers* complementares a NS5B, pistas F na Figura 3.3) dos oligonucleótidos desenhados tendo por base a sequência deste vírus. No entanto, a presença de um fragmento de DNA com tamanho de aproximadamente 200pb foi constatada aquando da utilização de *primers* complementares à sequência codificante de NS5B os quais permitem amplificar uma região conservada da região codificante desta proteína, presente no genoma de todos os flavivírus conhecidos (Parreira *et al.*, 2012) (Figura 3.3).



**Figura 3.3.** Observação dos produtos de PCR resultantes da amplificação de fragmentos de DNA correspondentes às regiões C, E, NS3hel e NS5B (extraído dos sedimentos celulares de C6/36 infectadas com o CFAV), utilizando os *primers* específicos e após a sua separação por electroforese em gel de agarose. Identificação das pistas: **(A)** células C6/36 não infectadas (controlo negativo). **(B)** amplificação da região codificante da proteína C; **(C)** amplificação da região codificante da proteína E; **(D)** amplificação da região codificante da proteína NS3hel e **(E)** amplificação da região codificante da proteína NS5B, a partir de DNA extraído dos sedimentos celulares de C6/36 infectadas com o CFAV. **(F)** amplificação de uma região altamente conservada na sequência codificante da proteína NS5B, presente no genoma de todos os flavivírus. **(M)** Marcador de massa molecular, representado à direita (*GeneRuler™ 100pb Plus DNA ladder*).

### 3.3. Amplificação dos fragmentos do genoma do CFAV a partir de DNA complementar (cDNA)

O RNA total extraído a partir de sobrenadante de cultura de células C6/36 infectadas com o CFAV foi utilizado para a produção de cDNA, através da sua síntese por transcrição reversa (RT) usando hexâmeros de sequência aleatória. Posteriormente, a presença de sequências complementares ao genoma do CFAV foi confirmada pela amplificação, por PCR, de parte das regiões codificantes da proteína C, E, NS3hel e NS5B, utilizando *primers* específicos (ver secção 2.4 do capítulo anterior). A análise dos produtos obtidos como resultado desta amplificação revelaram a presença, e tal como esperado, de fragmentos de DNA com tamanhos de aproximadamente 321pb, 1180pb, 1236pb e 2660pb, correspondente as regiões codificantes das proteínas C, E, NS3hel e NS5B (Figura 3.4), respectivamente.

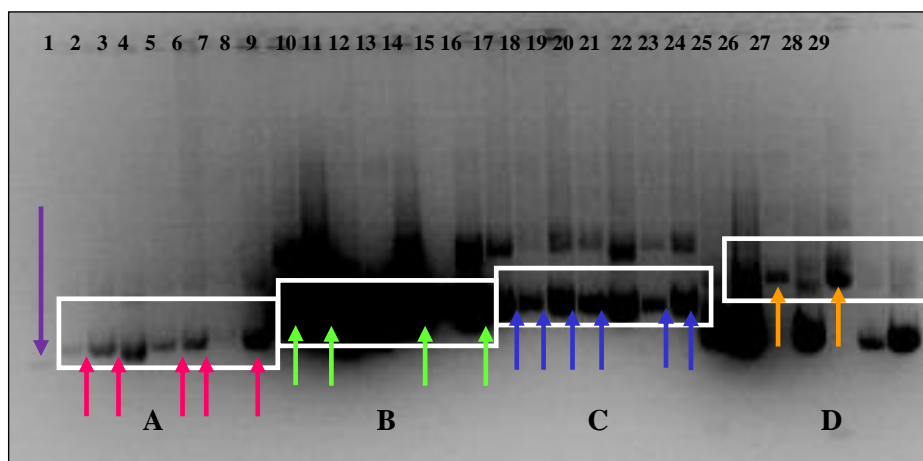


**Figura 3.4.** Separação por electroforese em gel de agarose dos fragmentos de DNA amplificados por PCR a partir de cDNA produzido por transcrição reversa (RT) a partir de RNA total extraído de sobrenadante de cultura de células C6/36 infectadas pelo CFAV. Identificação das pistas: **1-** Fragmento correspondente ao produto de amplificação referente a região codificante da proteína C. **2-** Fragmento correspondente ao produto de amplificação referente a região codificante da proteína E. **3-** Fragmento correspondente ao produto de amplificação referente a região codificante da proteína NS3hel. **4-** Fragmento correspondente ao produto de amplificação referente a região codificante da proteína NS5B. **(M)** Marcador de massa molecular (*GeneRuler™ 100pb Plus DNA ladder*). As setas assinalam a presença de fragmentos de DNA com os tamanhos esperados.

### 3.3.1. Clonagem molecular de fragmentos do genoma do CFAV no vector pJET1.2/*blunt*

Os quatro fragmentos de DNA correspondentes a parte das regiões codificantes das proteínas C, E, NS3<sub>hel</sub> e à totalidade da sequência codificante de NS5B de CFAV, uma vez amplificados por PCR, foram independentemente inseridos no vector pJET1.2/*blunt* usando *E. coli* NovaBlue como hospedeira de clonagem. O sistema de clonagem utilizado permitiu a selecção, *a priori*, dos clones bacterianos contendo plasmídeos recombinantes, uma vez que o vector utilizado possui um gene “letal” (codifica uma endonuclease de restrição de corte frequente) o qual, ao ser interrompido por ligação de um inserto de DNA no local de clonagem, permite que as células bacterianas com plasmídeos recombinante se propaguem. A razão pela qual foi realizado este passo, prendeu-se, exclusivamente, com o facto de nos permitir estabilizar as sequências codificantes das proteínas em estudo num vector de clonagem. Desta forma, tal procedimento permitiu a obtenção de uma boa quantidade de fragmentos de DNA que codificam as proteínas em estudo através da sua amplificação, posteriormente, por PCR, eliminando limitações impostas quanto à disponibilidade de material biológico de partida (cDNA).

Após expansão dos potenciais clones recombinantes em meio adequado (sete colónias para cada clone construído), e subsequente extracção do respectivo DNA plasmídico, as moléculas de DNA recombinante foram identificadas pela análise do seu padrão electroforético em gel de agarose. Esta análise foi baseada na comparação da migração de alíquotas de extractos de DNA plasmídicos, proveniente de diferentes clones, relativamente a um extracto de DNA plasmídico contendo moléculas de referências de tamanho semelhante ao vector utilizado (aproximadamente 3kpb). As potenciais moléculas recombinantes deveriam apresentar uma migração retardada relativamente às de referências (Figura 3.5.), como resultado da inserção de fragmentos de DNA exógeno (inserto), e do conseqüente aumento da respectiva massa molecular, quando comparada com o vector de clonagem.



**Figura 3.5.** Análise do padrão de migração, em gel de agarose, do DNA plasmídico de alguns dos clones bacterianos obtidos na sequência de clonagem de fragmentos de DNA referentes às regiões C, E, NS3hel e NS5B. A pista 1- DNA plasmídico correspondente ao vector *pBluescriptSKII*<sup>+</sup> (*Stratagene*). (A) DNA plasmídico correspondente a derivados do vector *pJET1.2/blunt*, contendo a região C de CFAV; (B) DNA plasmídico correspondente a derivados do vector *pJET1.2/blunt* potencialmente recombinantes, contendo a região E de CFAV; (C) DNA plasmídico correspondente a derivados do vector *pJET1.2/blunt*, contendo a região NS3hel de CFAV e (D) DNA plasmídico correspondente a derivados do vector *pJET1.2/blunt*, contendo a região NS5B de CFAV. As setas apontadas para cima indicam potenciais moléculas de DNA plasmídico recombinante. A seta apontada para baixo indica a forma circular, covalentemente fechada e super-enrolada do plasmídeo *pBluescriptSKII*<sup>+</sup> (*Stratagene*) com 3kpb.

Subsequentemente, a estrutura dos clones de DNA recombinante pretendidos foi verificada através de uma análise dos perfis electroforéticos dos fragmentos obtidos mediante a hidrólise dos plasmídeos em estudo com diferentes enzimas de restrição (ver secção 2.6.3), cujos locais de reconhecimento e corte flanqueiam o local de clonagem múltipla do vector utilizado (resultados não apresentados). Depois de confirmada a estabilização no vector *pJET1.2/blunt* das sequências codificantes das proteínas C, E, NS3hel e NS5B, foi possível obtê-las numa quantidade apreciável através da sua amplificação por PCR. Uma vez obtidos, esses fragmentos foram, então, digeridos com enzimas adequadas para que pudessem vir a ser clonados nos vectores de expressão (pET-28b e pET-29b).



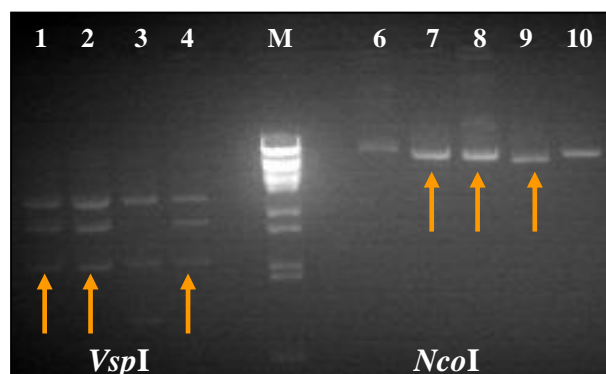
### 3.4. Clonagem das regiões codificantes das proteínas C, E, NS3hel e NS5B em vectores de expressão proteica (pET-28b e pET-29b)

Para a construção dos vectores de expressão contendo insertos correspondente às sequências codificantes das proteínas C, E, NS3hel e NS5B de CFAV, foi realizada a digestão dos plasmídeos em estudo com enzimas de restrição adequadas. Os fragmentos codificantes das proteínas C e E (produtos purificados) e o vector pET-28b foram duplamente digeridos com os enzimas de restrição *NcoI* e *NotI*. Quanto aos fragmentos codificantes das proteínas NS3hel e NS5B (produtos purificados), estes foram duplamente digeridos com os enzimas de restrição *VspI* e *NotI* e o vector pET-29b com os enzimas de restrição *NdeI* (que gera extremidades compatíveis com as geradas após clivagem com *VspI*) e *NotI* (procedimento descrito em 2.7.1.).

Após as reacções de digestão, procedeu-se à preparação de reacções de ligação dos insertos de DNA aos vectores de expressão, tal com descrito na secção 2.8. Em seguida, estas foram utilizadas na transformação de células competentes de *E. coli* NovaBlue. Foram seleccionadas colónias contendo moléculas de DNA potencialmente recombinantes e procedeu-se à extracção dos seus conteúdos em DNA plasmídico por lise alcalina, e usando uma estratégia analítica semelhante à descrita em 3.3.1, a qual nos permitiu seleccionar clones potencialmente recombinantes com base na migração retardada das moléculas de DNA plasmídico neles presentes relativamente às dos vectores de clonagem originais (resultados não apresentados).

A fase seguinte correspondeu à confirmação da presença dos insertos desejados no plasmídeo vector, a qual foi efectuada através de uma análise dos perfis electroforéticos dos fragmentos obtidos mediante a hidrólise dos plasmídeos em estudo com diferentes enzimas de restrição: *VspI* (pET-28b::C e pET-29b::NS3hel/NS5B), *XhoI* (pET-28b::E) e *NcoI* (pET-29b::NS3hel). A confirmação da presença dos fragmentos do DNA de interesse foi indicada pelo perfil de digestão enzimática obtido usando o enzima *VspI* e *NcoI*. Neste, foi revelado, como esperado, a presença de bandas com aproximadamente 2,76kpb, 2,17kpb e 1,43kpb correspondente ao plasmídeo pET-29b::NS3hel digerido com o enzima *VspI*. Este enzima corta cinco vezes o vector pET-29b, na posição 382, 1810, 1869, 4627 e 4816, e não clivando a sequência do inserto. Não foi possível visualizar as outras duas bandas por terem relativamente menores dimensões (59 e 189pb). Analisando o perfil enzimático originado pela digestão com o

enzima *NcoI* podemos visualizar a presença de uma banda com aproximadamente 7,0kpb, uma vez que este enzima corte apenas uma vez o vector (posição 212) (Figura 3.6.). Este procedimento foi equivalente para os restantes plasmídeos construídos.



**Figura 3.6.** Análise do perfil de digestão resultante da hidrólise, pelos enzimas *VspI* e *NcoI* de potenciais plasmídeos recombinantes contendo o inserto correspondente a região codificante da proteína NS3hel. (M) Marcador de tamanho molecular [ $\lambda$ DNA/*Eco9II* (*BstEII*)]. As setas indicam os fragmentos de DNA obtidos.

Após a verificação da identidade dos clones recombinantes através da análise do seu perfil de digestão enzimática, procedeu-se a extracção do DNA plasmídico, seguida da sua purificação, e à subsequente sequenciação usando, para tal, *primers* universais complementares a sequências localizadas nos limites do local de clonagem múltipla dos vectores pET utilizados. A sua sequenciação permitiu, essencialmente, constatar a inserção correcta dos insertos de interesse nos vectores de expressão, nomeadamente verificando a distância entre o promotor T7 e o codão AUG de iniciação da tradução, bem como da manutenção da grelha de leitura que permitiria a expressão das 4 proteínas do CFAV como extensões de 6 resíduos de histidinas nos seus extremos C-terminais (resultados não apresentados).

### 3.5. Expressão de proteínas recombinantes

#### 3.5.1. Efeito da indução da expressão de proteínas heterólogas sobre o crescimento de cultura de *E. coli* BL21(DE3)pLysS e Rosetta(DE3)pLysS

Um dos passos fundamentais para a expressão, com sucesso, de proteínas recombinantes em *E. coli* consiste na escolha, tão criteriosa quanto possível, da(s) estirpe(s) de expressão que virá(ão) a ser utilizada(s) para o efeito. Para este trabalho, foram selecionadas as estirpes a BL21(DE3)pLysS e Rosetta(DE3)pLysS de *E. coli*.

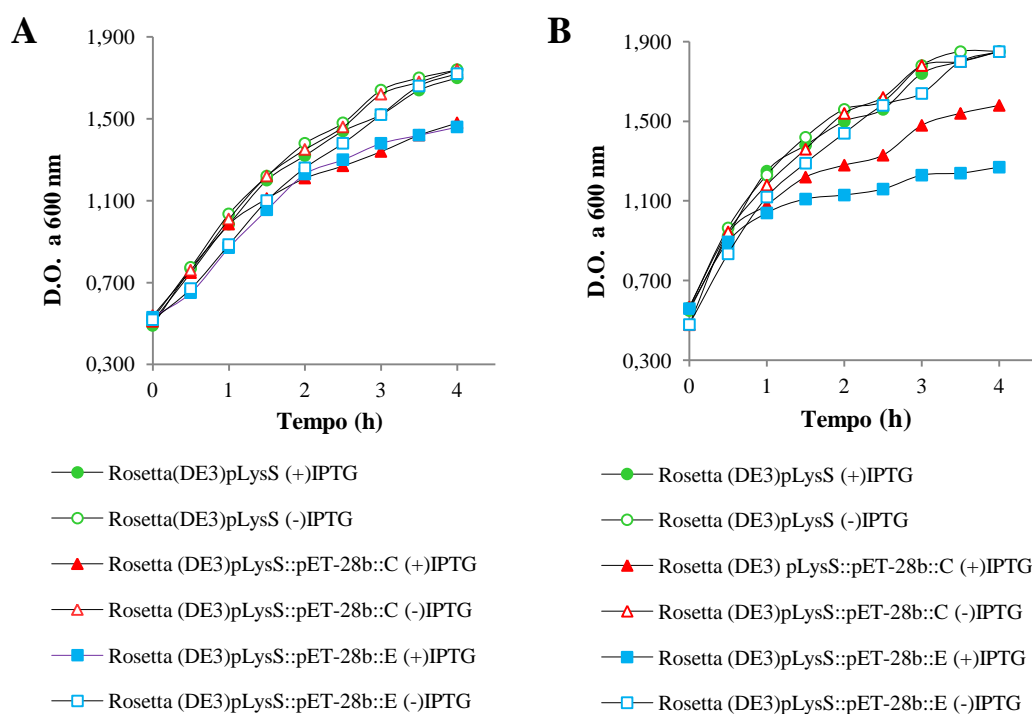
Frequentemente, a estirpe BL21(DE3)pLysS é usada na expressão de proteínas heterólogas em *E. coli* pois combina a ausência das proteases OmpT (localizada na membrana externa) e Lon (localizada no citoplasma), com a possibilidade de expressão proteica sob o controlo de um promotor de transcrição altamente específico (presente no genoma do fago T7). Este promotor é, por sua vez, controlado pela actividade da polimerase do fago T7, expressa a partir do genoma celular, onde a sua sequência codificante faz parte do fago lambda (DE3), e no qual se encontra sob o controlo transcricional de um promotor indutível pelo IPTG. A este, é aliada a manutenção de um nível baixo de expressão proteica [o qual se deve a alguma “permeabilidade” (*leakiness*) do sistema de expressão], através do controlo da actividade da polimerase de RNA fágica pelo seu inibidor natural, a lisozima do fago T7, expressa a partir do plasmídeo pLysS.

Houve necessidade de considerar, que os antígenos virais que se pretendiam expressar correspondem a proteínas codificadas por um vírus de células eucariontes. Na sequência codificante das mesmas constatou-se a presença de códons raramente utilizados pela maioria das estirpes de *E. coli* (Tabela 3.1), e que, por conseguinte, não deveriam ser reconhecidos de forma preferencial pelo *pool* de tRNAs acilados presente no citoplasma da estirpe de *E. coli* BL21(DE3)pLysS. Pelo contrário, a estirpe de *E. coli* Rosetta(DE3)pLysS transporta um derivado do plasmídeo pLysS (designado pRARE) onde se alia a sequência codificante da lisozima do fago T7 às que codificam tRNAs para alguns dos códons raramente utilizados por *E. coli* (AGG/Arg, AGA/Arg, AUA/Ile, CUA/Leu, CCC/Pro, e GGA/Gly). Devido ao eventual desajuste na utilização de códons foi feita uma análise, em paralelo, das diferenças de expressão nestas duas estirpes.

**Tabela 3.1.** Frequência com que aparecem os codões raros nas sequências das quatro regiões codificantes do genoma do CFAV (cápside, glicoproteína do invólucro, NS3hel e NS5B).

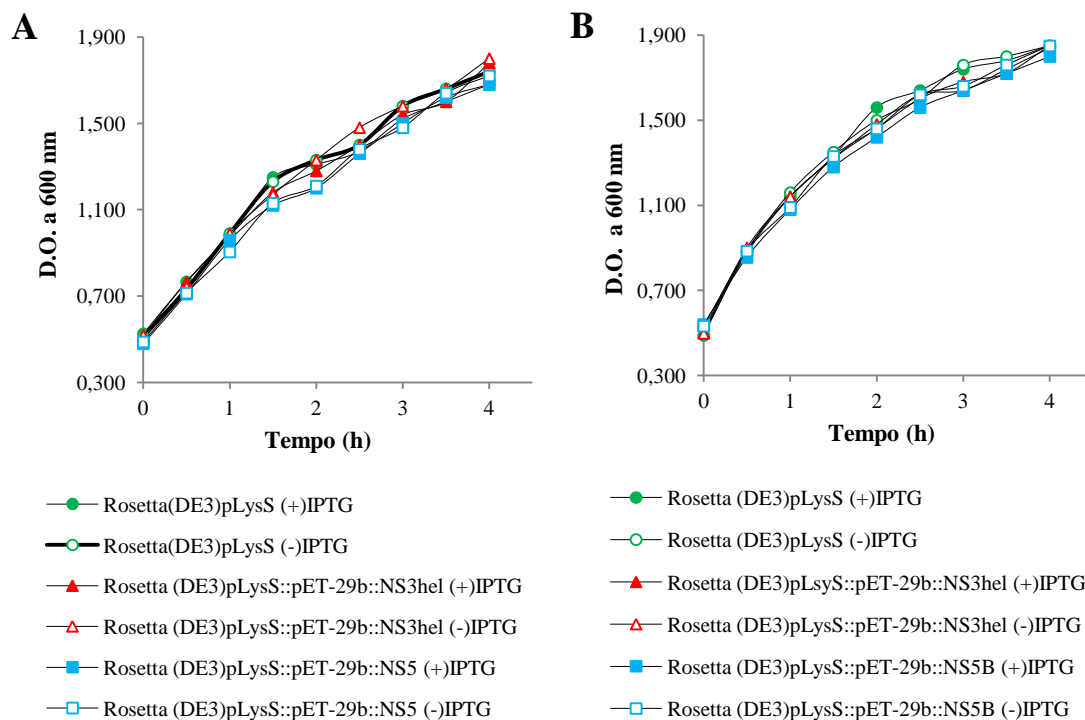
Proteína	Codões raros					
	AUA/Ile	AGG/Arg	AGA/Arg	CUA/Leu	GGA/Gly	CCC/Pro
<b>C</b>	0	4	4	1	8	2
<b>E</b>	4	4	2	5	14	1
<b>NS3hel</b>	4	6	4	2	11	4
<b>NS5B</b>	6	10	9	5	25	9

Para começar, os potenciais efeitos deletérios da expressão proteica sobre o crescimento bacteriano foram analisados seguindo o crescimento das mesmas (uma vez transformadas com plasmídeos de expressão) após a indução da expressão das proteínas pretendidas. Para tal, o crescimento de clones derivados das estirpes de *E. coli* BL21 (DE3)pLysS e de *E. coli* Rosetta(DE3)pLysS, transformadas, ou não, com os plasmídeos de expressão ou os seus derivados, foram seguidos através do registo da densidade óptica (600 nm) de culturas mantidas a 37°C nas quais, uma vez atingido o valor D.O.=0,5, a expressão proteica foi induzida com adição de IPTG para uma concentração final de 1mM (tal como sugerido pelo manual do sistema pET/Novagen, disponível em [http://pef.aibn.uq.edu.au/support/material/download/Novagen\\_pET\\_system\\_manual2003.pdf](http://pef.aibn.uq.edu.au/support/material/download/Novagen_pET_system_manual2003.pdf)). Estas culturas foram, então, mantidas a 30 e 37°C, com agitação, durante 4 horas. Os valores de D.O. obtidos pela monitorização do crescimento bacteriano com leituras feitas a intervalos regulares de 30 minutos foram utilizados para a construção de curvas de crescimento para cada uma das estirpes. Ao contrário do que foi registado para os derivados de *E. coli* BL21(DE3)pLysS contendo os 4 plasmídeos de expressão construídos e nas quais, a indução de expressão de cada uma das proteínas pretendidas, quer a 30 ou 37°C, não teve qualquer impacto no crescimento das mesmas (resultados não apresentados). Os registos relativos ao crescimento dos derivados de *E. coli* Rosetta(DE3)pLysS estão graficamente representados nas Figuras 3.7. e 3.8.



**Figura 3.7.** Curvas de crescimento em meio LB. (A) Figura representa as curvas de crescimento de culturas de *E.coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-28b::C e pET-28b::E, não induzidas e induzidas a uma densidade óptica (D.O.  $_{600\text{ nm}}$ ) de 0.5 com concentração de 1 mM de IPTG e incubadas a 30°C. (B) Figura representa as curvas de crescimento de culturas de *E.coli* Rosetta (DE3)pLysS não transformada e transformada com os plasmídeos pET-28b::C e pET-28b::E, não induzidas e induzidas a uma densidade óptica (D.O.  $_{600\text{ nm}}$ ) de 0.5 com concentração de 1 mM de IPTG e incubadas a 37°C. Nos painéis A e B, o tempo é indicado em horas após o início da indução.

Como se pode verificar pela análise da Figura 3.7.B, à temperatura de 37°C as culturas de *E. coli* Rosetta(DE3)pLysS uma vez induzida a expressão das proteínas C e E de CFAV, apresentam uma redução na taxa de crescimento, quando comparadas com as culturas transformadas não induzidas. Esta redução foi, especialmente, evidente no caso proteína viral E. Este dado sugere que a desaceleração no crescimento bacteriano (relativamente às estirpes onde a expressão proteica não foi induzida) possa resultar de um qualquer efeito tóxico associado à produção das proteínas heterólogas em questão ou, em alternativa, da mobilização da maquinaria celular para a produção das proteínas do CFAV, reduzindo, assim, o crescimento bacteriano. Por outro lado, as mesmas culturas, quando induzidas e incubadas à temperatura de 30°C, apresentam uma ligeira redução na taxa de crescimento, quando comparadas com as culturas não induzidas.



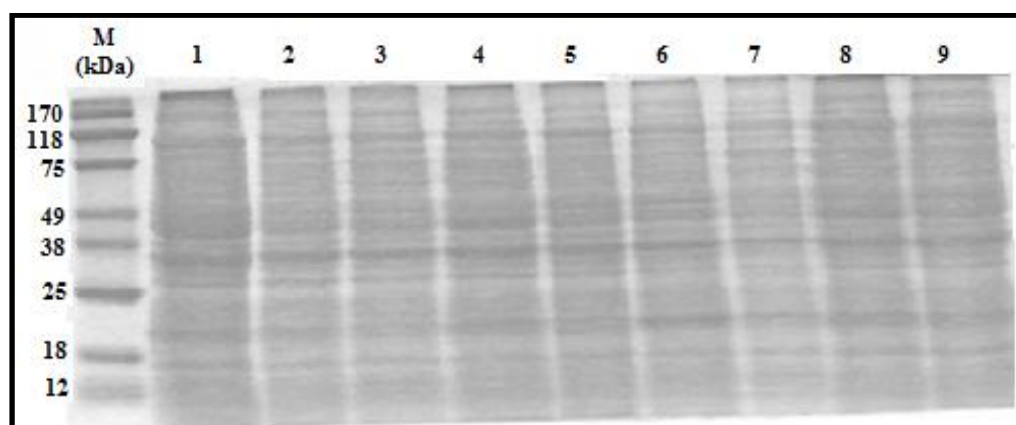
**Figura 3.8.** Curvas de crescimento em meio LB. (A) Figura representa as curvas de crescimento de culturas de *E.coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-29b::NS3hel e pET-29b::NS5B, não induzidas e induzidas a uma densidade óptica (D.O.  $_{600\text{ nm}}$ ) de 0.5 com concentração de 1 mM de IPTG, e incubadas a 30°C. (B) Figura representa as curvas de crescimento de culturas de *E.coli* Rosetta (DE3)pLysS não transformada e transformada com os plasmídeos pET-29b::NS3hel e pET-29b::NS5B, não induzidas e induzidas a uma densidade óptica (D.O.  $_{600\text{ nm}}$ ) de 0.5 com concentração de 1 mM de IPTG, e incubadas a 37°C. Nos painéis A e B, o tempo é indicado em horas após o início da indução.

Como pode ser verificado pela figura acima, as culturas de *E. coli* Rosetta (DE3)pLysS transformadas com os plasmídeos pET-29b::NS3hel e pET-29b::NS5B induzidas e incubadas às temperaturas 30 e 37°C tiveram um comportamento (taxa de crescimento) similar, quando comparadas com as culturas transformadas (não induzidas) e com as culturas não transformadas (não induzidas e induzidas). Este dado indica que a produção destas proteínas nesta estirpe não parece exercer qualquer efeito de toxicidade com impactos directos sobre o metabolismo celular ou promover, por mobilização de recursos metabólicos, uma desaceleração no crescimento das mesmas.

Enquanto que o registo, aquando da indução da expressão de proteínas heterólogas em *E. coli*, de um impacto, mais ou menos significativo, de alterações no crescimento celular, pode ser indicativa de que estas estão, de facto, a ser produzidas *in*

*vivo*, o contrário não é, necessariamente verdadeiro. Ou seja, a ausência de um efeito da indução da expressão de proteínas sobre o crescimento de uma estirpe bacteriana pode, simplesmente, traduzir uma ausência de efeitos nefastos da sua produção e acumulação sobre o metabolismo celular, e não ausência da sua síntese. Por esta razão, em paralelo com a monitorização do impacto da presumível síntese de proteínas codificadas pelo CFAV em *E. coli*, procedeu-se à recolha de alíquotas durante a monitorização do crescimento bacteriano, de forma a constatar a expressão das proteínas heterólogas pretendidas. Após a sua recolha, as células contidas nos diferentes sedimentos celulares foram ressuspensas em tampão de amostra e desnaturadas por calor, conforme o procedimento descrito na secção 2.9.4. Os extractos proteicos totais obtidos a partir de derivados de *E. coli* BL21(DE3)pLysS (transformada, ou não, com os plasmídeos de expressão) foram analisados por SDS-PAGE. Esta análise foi baseada na comparação de bandas proteicas identificadas nos extractos analisados relativamente a extractos derivados de *E. coli* BL21(DE3)pLysS transformadas, apenas, com os vectores pET-28b e pET-29b, e a um marcador de massa molecular contendo bandas proteicas de referência (massa molecular conhecida).

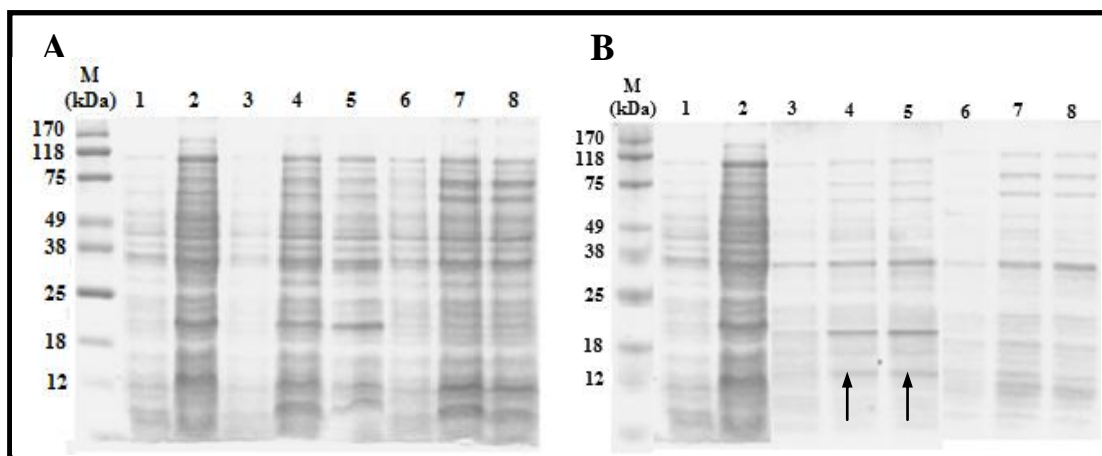
Através desta análise foi possível constatar, a aparente ausência, 4 horas após indução a 30 e a 37°C, das proteínas desejadas (Figura 3.9), em extractos proteicos totais obtidos a partir de derivados de *E. coli* BL21(DE3)pLysS. Estes dados indicam que, muito provavelmente, a não observação de qualquer tipo de impacto da indução da expressão de proteínas heterólogas no crescimento destas estirpes se deve à ausência de expressão das proteínas C, E, NS3hel e NS5B, ou à sua produção em quantidades tão diminutas que relem a sua detecção para a utilização de métodos analíticos mais sensíveis (exemplo: detecção com anticorpos ou marcação radioactiva).



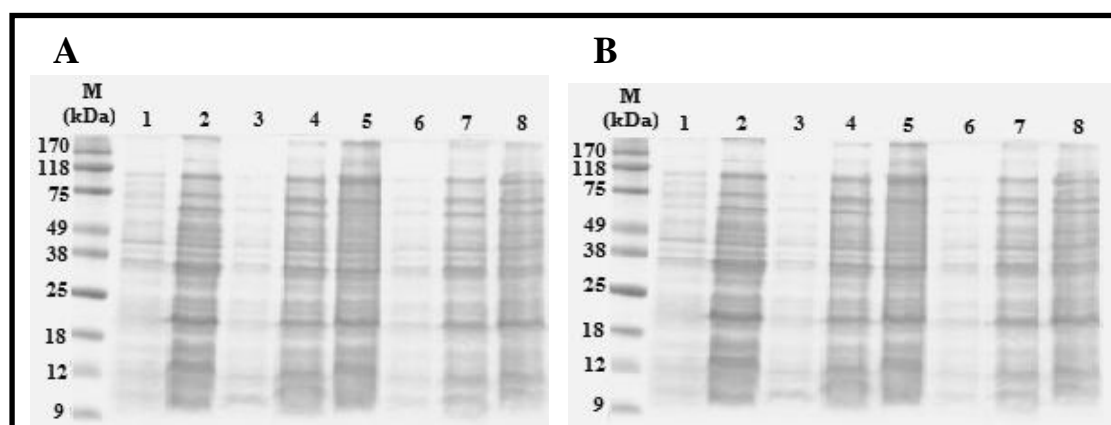
**Figura 3.9.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais derivados de culturas de *E. coli* BL12(DE3)pLysS não transformada e transformada com os plasmídeos construídos, 4 horas após indução com 1 mM de IPTG a 30 e 37°C. Identificação das pistas: M - marcador de massa molecular (Lonza, EUA); 1-Controlo (estirpe sem o vector de expressão); 2, 3, 4 e 5- extractos proteicos totais derivados de culturas de *E. coli* BL21(DE3)pLysS transformada com os plasmídeos pET-28b::C, pET-28b::E, pET-29b::NS3hel e pET-29b::NS5B, induzidos a 30°C; 6, 7, 8 e 9- extractos proteicos totais derivados de culturas de *E. coli* BL21(DE3)pLysS transformada com os plasmídeos pET-28b::C, pET-28b::E, pET-29b::NS3hel e pET-29b::NS5B, induzidos a 37°C.

De forma a constatar se a expectável presença em *E. coli* Rosetta(DE3)pLysS de tRNAs acilados não existente em *E. coli* BL21(DE3)pLysS se traduziu na produção de uma quantidade significativamente diferente das diferentes proteínas do CFAV, extractos proteicos totais de derivados de culturas de *E. coli* Rosetta(DE3)pLysS foram também analisados por SDS-PAGE. Esta análise permitiu, de facto, confirmar a presença de uma banda com aproximadamente 12kDa, correspondente à massa estimada da proteína C, nos extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS transformada com o plasmídeo pET-28b::C, depois das culturas bacterianas terem sido induzidas a temperatura de 37°C com 1mM de IPTG durante 4 horas. No entanto, o mesmo não se verificou quando as culturas bacterianas foram induzidas a temperatura de 30°C com 1mM de IPTG durante 4 horas. Por outro lado, nem sempre foi possível visualizar, de forma clara, as bandas dos extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS transformada com os plasmídeos pET-28b::E, pET-29b::NS3hel e pET-29b::NS5B, após a indução da expressão proteica a temperatura de 30 e 37°C com 1mM de IPTG durante 4 horas. Tal facto, pode ocorrer quando há sobreposição da banda correspondente a proteína recombinante com outra banda de *E. coli* Rosetta(DE3)pLysS. Além disso, as bandas esperadas correspondentes as proteínas C, E, NS3hel e NS5B não são visualizadas em extractos proteicos preparados a partir de culturas que não foram sujeitas a tratamento com IPTG, dando a indicação de que não ocorre “expressão basal” considerável (Figura 3.10. e Figura 3.11.).





**Figura 3.10.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos construídos, 4 horas após indução com 1 mM de IPTG a 30 °C e 37°C. (A) A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-28b::C e pET-28b::E, após indução com 1mM de IPTG a 30°C. Identificação das pistas: M - marcador de massa molecular (Lonza, EUA); 1e 2-não transformada (indução t=0h e t=4h); 3, 4 e 5- transformada com o plasmídeo pET-28b::C (indução t=0h, t=2h e t=4h); e 6, 7 e 8- transformada com o plasmídeo pET-28b::E (indução t=0h, t=2h e t=4h). (B) A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos plasmídeo pET-28b::C e pET-28b::E, após indução com 1mM IPTG a 37°C. Identificação das pistas: M - marcador de massa molecular (Lonza, EUA); 1e 2-não transformada (indução t=0h e t=4h); 3, 4 e 5- transformada com o plasmídeo pET-28b::C (indução t=0h, t=2h e t=4h); e 6, 7 e 8- transformada com o plasmídeo pET-28b::E (indução t=0h, t=2h e t=4h). As setas indicam as bandas presumivelmente correspondentes à proteína C.



**Figura 3.11.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos construídos, 4 horas após indução com 1 mM de IPTG a 30 °C e 37°C. (A) A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS

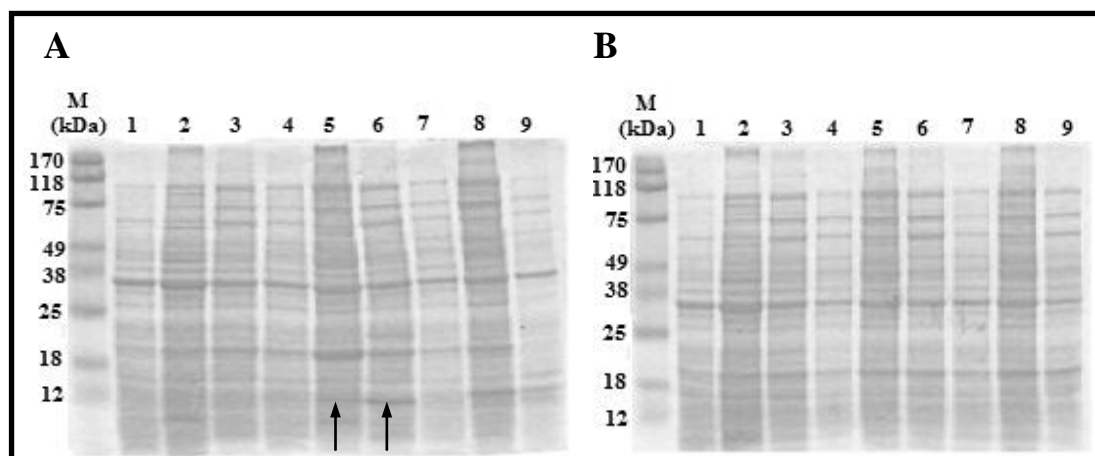
não transformada e transformada com os plasmídeos pET-29b::NS3hel e pET-29b::NS5B, após indução com 1mM de IPTG a 30°C. Identificação das pistas: M - marcador de massa molecular (Lonza, EUA); 1e 2-não transformada (indução t=0h e t=4h); 3, 4 e 5- transformada com o plasmídeo pET-29b::NS3hel (indução t=0h, t=2h e t=4h); e 6, 7 e 8- transformada com o plasmídeo pET-29b::NS5B (indução t=0h, t=2h e t=4h). **(B)** A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos plasmídeo pET-29b::NS3hel e pET-29b::NS5B, após indução com 1mM IPTG a 37°C. Identificação das pistas: M - marcador de massa molecular (Lonza, EUA); 1e 2-não transformada (indução t=0h e t=4h); 3, 4 e 5- transformada com o plasmídeo pET-29b::NS3hel (indução t=0h, t=2h e t=4h); e 6, 7 e 8- transformada com o plasmídeo pET-29b::NS5B (indução t=0h, t=2h e t=4h).

Após a avaliação da indução da expressão através da monitorização do crescimento bacteriano, procedeu-se com a optimização de dois parâmetros distintos, utilizando apenas a estirpe de *E. coli* Rosetta(DE3)pLysS, por razões mencionadas acima.

### **3.5.2. Efeito da temperatura e concentração de indutor na expressão de proteína**

A verificação dos níveis de expressão proteica foi levada a cabo através da avaliação que nela têm dois parâmetros distintos: a temperatura de crescimento bacteriano e a concentração do indutor (IPTG). Estes parâmetros podem influenciar, independentemente, o crescimento celular e a produção de proteína. Foram testadas três temperaturas diferentes: 15, 25 e 37°C, sendo que a indução da produção de proteínas heterólogas foi induzida pela adição de 1mM de IPTG durante aproximadamente 16 horas. Findo este período de tempo procedeu-se à recolhida de alíquotas que, posteriormente, foram ressuspensas em tampão de amostra para posterior análise por SDS-PAGE (procedimento descrito em 2.9.4 capítulo anterior). Através desta análise foi possível constatar a expressão da proteína C, revelada pela identificação no extracto proteico total derivado de uma cultura mantida a 37°C e induzida com IPTG de uma banda com aproximadamente 12kDa. Esta mesma banda foi visualizada no extracto proteico total derivado de uma cultura mantida a 25°C e induzida com IPTG, mas indicando que o nível de expressão foi mais baixo quando comparado com o extracto proteico derivado de uma cultura mantida a 37°C (e induzida com IPTG). O mesmo não aconteceu no extracto proteico derivado de uma cultura mantida a 15°C e induzida com

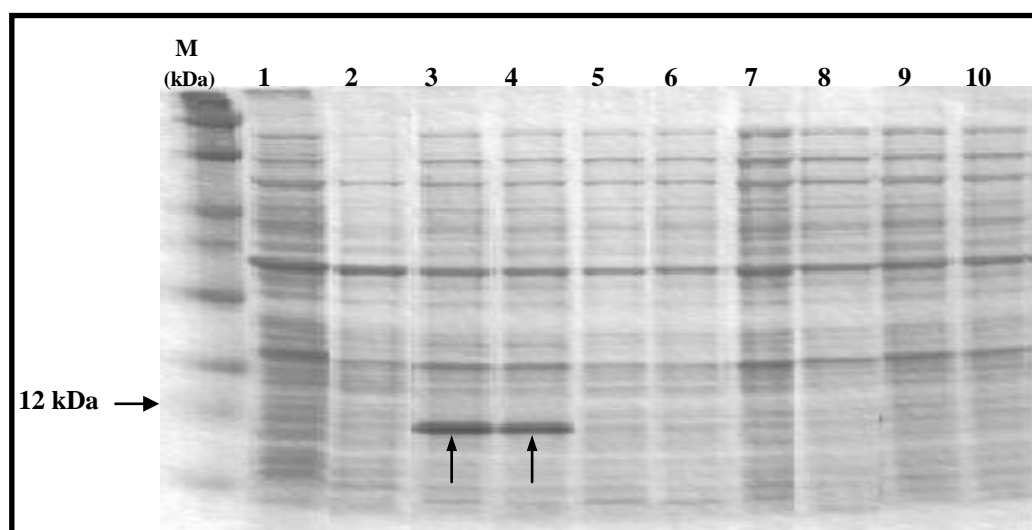
IPTG, em que não foi possível visualizar a mesma banda. Esta mesma banda não é foi visualizada, como esperado, nos extractos proteicos derivados das culturas mantidas a 15, 25 e 37°C e induzidas com IPTG, correspondentes a *E. coli* Rosetta(DE3)pLysS não transformada com o vector pET-28b (Figura 3.12.A). No entanto, não foi possível visualizar de forma clara bandas correspondentes as esperadas proteínas E, NS3hel e NS5B nos extractos proteicos totais das estirpes onde deveriam estar presentes (Figura 3.12.A e Figura 3.12.B).



**Figura 3.12.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) dos extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos construídos, após a indução com 1mM de IPTG a 15, 25 e 37°C durante a noite. **(A)** A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos pET-28b::C e pET-28b::E. Identificação das pistas: M- marcador de massa molecular (Lonza, EUA); 1, 2 e 3- não transformada (indução a 15, 25 e 37°C); 4, 5 e 6- transformada com o plasmídeo pET-28b::C (indução a 15, 25 e 37°C); e 7, 8 e 9- transformada com o plasmídeo pET-28b::E (indução a 15, 25 e 37°C). **(B)** A figura representa os extractos proteicos totais derivados de culturas de *E. coli* Rosetta(DE3)pLysS não transformada e transformada com os plasmídeos plasmídeo pET-29b::NS3hel e pET-29b::NS5B. Identificação das pistas: M- marcador de massa molecular (Lonza, EUA); 1, 2 e 3- não transformada (indução a 15, 25 e 37°C); 4, 5 e 6- transformada com o plasmídeo pET-29b::NS3hel (indução a 15, 25 e 37°C); e 7, 8 e 9- transformada com o plasmídeo pET-29b::NS5B (indução a 15, 25 e 37°C). As setas indicam as bandas presumivelmente correspondentes à proteína C.

Outro parâmetro avaliado foi a concentração do indutor (IPTG), tendo sido analisada a expressão de proteínas durante 4 horas a temperatura de 37°C (procedimento descrito na secção 2.9.2) testando duas concentrações (0,4 e 1 mM) desta substância. Os

extractos proteicos totais obtidos a partir de derivados de *E. coli* Rosetta(DE3)pLysS (transformada com os plasmídeos pET-28b::C/E e pET-29b::NS3hel/NS5B) foram analisados por SDS-PAGE. Esta análise revelou que os níveis de expressão proteica resultantes da utilização de duas concentrações distintas de IPTG foram, essencialmente, equivalentes, tendo sido possível visualizar uma banda com aproximadamente 12kDa, correspondente à massa estimada para a proteína C, tanto nos extractos proteicos totais derivados de culturas induzidas com 0,4 mM de IPTG como nos extractos proteicos totais derivados de culturas induzidas com 1 mM de IPTG. No entanto, não foi possível visualizar bandas correspondentes às proteínas E, NS3hel e NS5B, tanto nos extractos proteicos totais derivados de culturas induzidas com 0,4 mM de IPTG como nos extractos proteicos totais derivados de culturas induzidas com 1 mM de IPTG (Figura 3.13.).



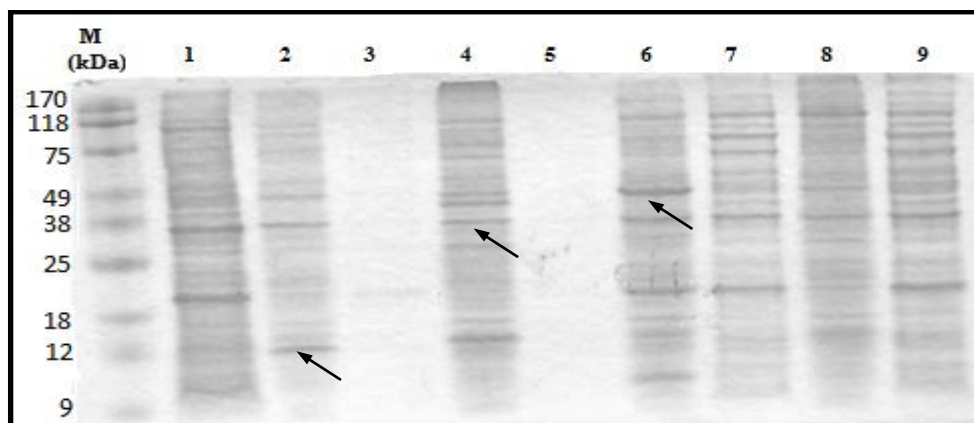
**Figura 3.13.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) dos extractos proteicos totais derivados de culturas de *E. coli* Rosetta (DE3)pLysS não transformada e transformada com os plasmídeos construídos, após indução com 0,4 e 1 mM de IPTG a 37°C durante 4 horas. Identificação das pistas: M- marcador de massa molecular (Lonza, EUA); 1- controlo (estirpe sem vector de expressão) onde a expressão proteica não foi induzida; 2- controlo (estirpe sem vector de expressão) onde a expressão proteica foi induzida com 1mM de IPTG; 3 e 4- transformada com o plasmídeo pET-28b::C (indução com 0,4 e 1 mM de IPTG); 5 e 6- transformada com o plasmídeo pET-28b::E (indução com 0,4 e 1 mM de IPTG); 7 e 8- transformada com o plasmídeo pET-29b::NS3hel

(indução com 0,4 e 1 mM de IPTG); 9 e 10- transformada com o plasmídeo pET-29b::NS5B (indução com 0,4 e 1 mM de IPTG). As setas indicam as bandas, provavelmente, correspondentes a proteína C.

Após a verificação dos dois parâmetros optou-se por, nas fases seguintes, induzir a expressão das proteínas em estudo com 1 mM de IPTG durante 4 horas a 37°C. (condições de referência, tal como sugerido em [http://pef.aibn.uq.edu.au/support/material/download/Novagen\\_pET\\_system\\_manual2003.pdf](http://pef.aibn.uq.edu.au/support/material/download/Novagen_pET_system_manual2003.pdf)).

### **3.5.3. Verificação da solubilidade de proteínas heterólogas**

A verificação da solubilidade das proteínas recombinantes expressas em *E. coli* foi feita analisando a fracção proteica de sedimentos celulares de culturas induzidas com 1mM de IPTG e incubadas durante 4 horas a 37°C. Estes sedimentos celulares foram ressuspensos em tampão de lise (ver na secção 2.9.3.) e as células desintegradas por sonicação (ciclos com 60% de amplitude máxima, com um rácio de 0,5 entre as emissões sonoras e as pausa cíclicas, durante um 1 a 10 minutos). Após uma centrifugação, foram obtidas duas fracções: uma solúvel (correspondendo ao sobrenadante) e uma insolúvel (correspondendo ao sedimento), tendo estas duas fracções sido, de seguida, analisadas por electroforese em gel de poliacrilamida (Tricina-SDS-PAGE). Através desta análise foi possível visualizar bandas correspondendo a proteínas com aproximadamente 12, 43 e 49kDa, supostamente correspondentes às proteínas C, E e NS3hel, na fracção insolúvel, dando a indicação de que nas estirpes onde a sua produção foi induzida tenha ocorrido a formação de que corpos de inclusão. Quanto as pistas correspondentes as fracções solúveis referentes as proteínas C e E, revelam a, quase, ausência de bandas proteicas e, tal facto, muito provavelmente, deve-se à lise incompleta das células durante a sonicação. Por outro lado, não foi possível visualizar a banda correspondente a proteína NS5B, tanto na fracção solúvel como na fracção insolúvel. Este facto, muito provavelmente, deve-se à sua produção em quantidades diminutas, que relegam a sua detecção para a utilização de métodos analíticos mais sensíveis (por exemplo: detecção com anticorpos ou marcação radioactiva) (Figura 3.14.).



**Figura 3.14.** Análise do padrão electroforético em gel de poliacrilamida (Tricina-SDS-PAGE) de alíquotas das fracções solúveis (Fs) e insolúveis (Fi) obtidas a partir de extractos proteicos totais derivados de culturas de *E. coli* Rosetta (DE3)pLysS não transformada e transformada com os plasmídeos construídos, onde foi induzida a produção da proteína C, E, NS3hel e NS5B. Identificação das pistas: M- marcador de massa molecular (Lonza, EUA); 1- extracto proteico total derivados de culturas de *E. coli* Rosetta (DE3)pLysS não transformada; 2 e 3- fracção insolúvel e solúvel correspondente a proteína C; 4 e 5- fracção insolúvel e solúvel correspondente a proteína E; 6 e 7- fracção insolúvel e solúvel correspondente a proteína NS3hel; e 8 e 9- fracção insolúvel e solúvel correspondente a proteína NS5B; As setas indicam, provavelmente, as proteínas de interesse.

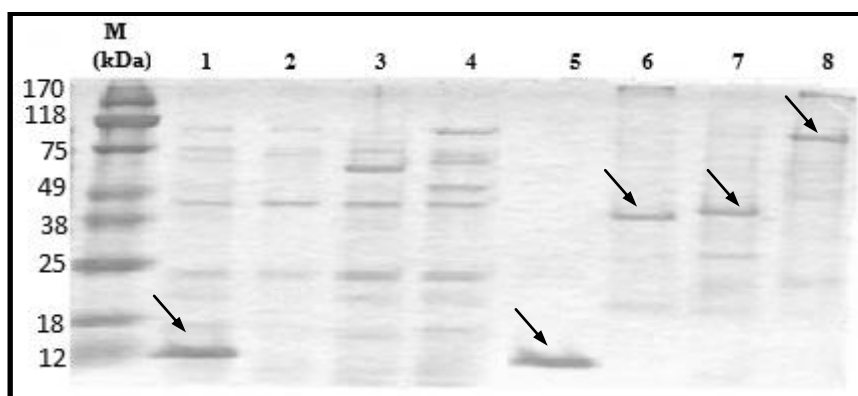
#### 3.5.4. Purificação de proteínas recombinantes com “caudas” de hexahistidina

Para a confirmação da identidade das proteínas C, E, NS3hel e NS5B expressas em *E. coli* como produtos recombinantes fundidos a uma “cauda” de histidina, prosseguiu-se com a purificação das mesmas por cromatografia de afinidade com iões níquel. A purificação foi realizada forçando (por centrifugação) a passagem de um extracto proteico através de uma coluna de cromatografia empacotada com resina contendo iões níquel imobilizados. Esta técnica foi realizada em condições desnaturantes e em condições nativas (procedimentos experimentais descritos em 2. 10. 1. e 2.10.2.).

A desnaturação proteica foi assegurada pela presença de uma elevada concentração (8 M) de ureia, uma vez que este composto, exercendo uma acção caotrópica, vai desnaturando e solubilizando as diferentes proteínas presentes num extracto. As soluções de ligação e lavagem das colunas onde proteínas com caudas de histidinas são imobilizadas incluíam na sua composição imidazol. Esta molécula

apresenta uma estrutura química em anel, com afinidade para os iões níquel, com a finalidade de reduzir a formação de ligações não específicas entre estes iões e eventuais proteínas contaminantes com menor afinidade para a matriz. O imidazol, em elevada concentração, pode ainda ser usado para a eluição da proteína que se pretende purificar. Neste trabalho optou-se, pela eluição com três concentrações diferentes de imidazol, 100, 200 e 500 mM. Realizaram-se entre quatro a seis passos de lavagens, com 600 µl cada, e três passos de eluição com 300 µl cada. Posteriormente, realizou-se a análise dos produtos obtidos (eluatos) por electroforese em gel de poliacrilamida (Tricina-SDS-PAGE).

A análise dos eluatos obtidos na sequência da purificação de proteínas em condições desnaturantes permitiu a detecção das proteínas com massas moleculares compatíveis com as calculadas (teoricamente) para C, E, NS3hel e NS5B a partir do primeiro eluato. Neste, estas apresentavam um nível de purificação considerável, revelado pela presença de bandas proteicas bem evidentes com massas moleculares esperadas de aproximadamente 12, 43, 49 e 100kDa (pista 5, 6, 7 e 8). No entanto, em quase todos eles constatou-se ainda a presença de uma quantidade variável de proteínas contaminantes (nomeadamente nos extractos contendo as proteínas E, NS3hel e NS5B) que não foram removidas das colunas de cromatografia pela solução de lavagem utilizada. Uma forma de contornar este problema consistiu na adição 30 mM de imidazol ao tampão de ligação das proteínas à coluna de cromatografia, com o objectivo de diminuir a ligação inespecífica de outras proteínas à mesma. No entanto, mesmo nestas condições, os eluatos revelaram, repetidamente, uma pequena quantidade de proteínas contaminantes. Ainda assim, foi possível obter a proteína C com um nível de purificação considerável (Figura 3.15.).



**Figura 3.15.** Resultados da purificação por cromatografia de afinidade com iões  $\text{Ni}^{2+}$  das proteínas C, E, NS3hel e NS5B a partir de lisados bacterianos. Identificação das pistas: M- marcador de massa molecular (Lonza, EUA); 1, 2, 3 e 4- extractos proteicos purificados em condições nativa correspondentes as proteínas C, E, NS3hel e NS5B; 5, 6, 7 e 8- extractos proteicos purificados em condições desnaturantes correspondentes as proteínas C, E, NS3hel e NS5B. As setas indicam as proteínas recombinantes purificadas.

Como se pode verificar pela análise da figura 3.15., a eluição das proteínas C, E, NS3hel e NS5B purificadas em condições nativas (pista 1, 2, 3 e 4), revelou a presença de uma banda correspondente a uma proteína com uma massa compatível com a estimada (teoricamente) para a proteína C. No entanto, não só a presença das proteínas E, NS3hel e NS5B não foi evidente como também se verificou a presença de uma grande quantidade de contaminantes nos eluatos obtidos, proteínas estas que permaneceram fixadas às colunas cromatográficas apesar das sucessivas lavagens das mesmas. A não fixação, em condições nativas, das proteínas E, NS3hel e NS5B às colunas de cromatografia poderá dever-se ao facto de que, durante a sua produção, estas possam ter adquirido uma conformação, de tal forma que a “cauda” de histidina pode não ter ficado exposta na superfície proteica, impedindo, então, a sua ligação à resina. Em contrapartida, a solubilização dos corpos de inclusão com 8 M de ureia terá promovido a desnaturação de todas (ou pelo menos a maioria) das proteínas nos extractos produzidos, de forma que a “exposição” das “caudas” de histidina, terá viabilizado a sua ligação aos iões níquel da coluna.

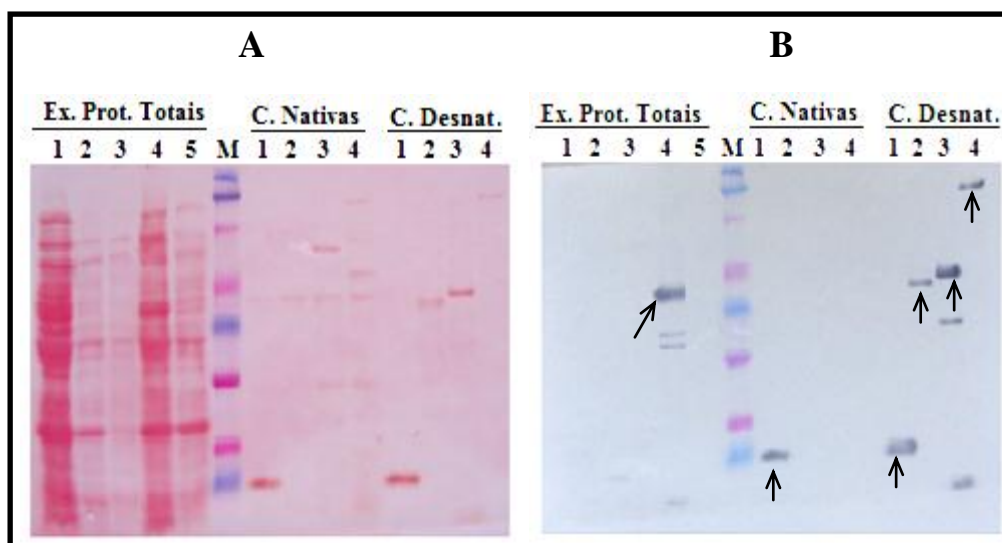
Uma forma de confirmar a autenticidade das proteínas produzidas e purificadas com “caudas” de histidina, e que têm a massa molecular esperada, é a sua detecção utilizando métodos analíticos tais como a sua detecção com anticorpos (*Western Blot*).

### **3.5.5. Detecção de proteínas por *Western Blot***

Com objectivo de confirmar a presença de “caudas” de histidina nas proteínas purificadas em condições nativas e desnaturantes, procedeu-se à detecção das mesmas por *Western Blot*, usando, para tal, um anticorpo anti-histidina, conforme procedimento descrito na secção 2.11. do capítulo anterior. Os resultados de imunodeteção, referentes às proteínas purificadas, em condições desnaturantes, a partir dos extractos proteicos totais das diferentes estirpes derivadas de *E. coli* Rosetta(DE3)pLysS



expressando proteínas do CFAV, permitiram confirmar, de forma específica, a presença de domínios ricos em histidinas em quatro proteínas com aproximadamente 12, 43, 49 e 100kDa. Tal como esperado, estas deverão corresponder as proteínas C, E, NS3hel e NS5B (pista 1, 2, 3 e 4). Quanto aos extractos proteicos obtidos em condições nativas, foi possível detectar, apenas, a proteínas C (pista 1, Figura 3.16.B). Por outro lado, os resultados de imunodeteccção, referentes aos extractos proteicos totais obtidos das diferentes estirpes derivadas de *E. coli* BL21(DE3)pLysS expressando proteínas do CFAV, revelaram a ausência de domínios ricos em histidinas em proteínas, que deveriam corresponder as proteínas C, E e NS5B (pista 2, 3 e 5). No entanto, na pista 4 em que foi possível confirmar a presença de um domínio rico em histidina numa proteína com aproximadamente 49kDa, a qual, deverá corresponder a proteína NS3hel (Figura 3.16.B).



**Figura 3.16.** Detecção específica das proteínas recombinantes. **(A)** A figura representa o gel corado com solução de *Rouge Ponceau* antes de realizar-se o “*Western Blot*”. Da esquerda para direita: extractos proteicos totais: 1- *E. coli* BL21(DE3)pLysS não transformada (controlo); 2, 3, 4 e 5- *E. coli* BL21(DE3)pLysS transformada com os plasmídeos pET-28b::C/E, e pET-29b::NS3hel/NS5B. (M) marcador de massa molecular (Lonza, EUA). Extractos purificados em condições nativas e desnaturantes, correspondentes a *E. coli* Rosetta(DE3)pLysS transformada com os plasmídeos pET-28b::C/E e pET-29b::NS3hel/NS5B: 1- C, 2- E, 3- NS3hel e 4- NS5B. **(B)** A figura representa o gel de “*Western Blot*”. Da esquerda para direita: extractos proteicos totais: 1- *E. coli* BL21(DE3)pLysS não transformada; 2, 3, 4 e 5- *E. coli* BL21(DE3)pLysS transformada com os vectores pET-28b::C/E e pET-29b::NS3hel/NS5B. (M)

marcador de massa molecular (Lonza, EUA). Extractos purificados em condições nativas e desnaturantes, correspondente a *E. coli* Rosetta(DE3)pLysS transformada com os vectores pET-28b::C/E e pET-29b::NS3hel/NS5B: 1- C, 2- E, 3- NS3hel e 4- NS5B. As setas indicam as proteínas recombinantes detectadas.

Abaixo da banda da proteína de interesse (pista 4 a esquerda e pista 3 à direita, Figura 3.16.B) reconhecida pelo anticorpo anti-histidina, registou-se há presença de outras bandas correspondendo a proteínas com massas inferior à esperada, que igualmente reagiram com o anticorpo anti-histidina. Estas devem corresponder a produtos resultantes da degradação de NS3hel.

## ***4. DISCUSSÃO DE RESULTADOS***

---

#### 4. DISCUSSÃO DE RESULTADOS

O *Cell Fusing Agent Virus* (CFAV) foi o primeiro “flavivírus específico de insectos” (ISF) a ser descrito (Stollar e Thomas, 1975) e, tal como os demais ISF, é expectável que, na Natureza, esteja exclusivamente adaptado ao seu hospedeiro natural (*Aedes aegypti*). Ainda que, *in vitro*, a sua replicação ocorra em células de *Ae. albopictus* (como são as que compõem a linha celular designada C6/36), não replica em células de vertebrados (Cammisa-Parks *et al.*, 1992).

Apesar de ter sido identificado há mais de três décadas (1975), a verdade é que muito pouco (ou nada) se conhece sobre a sua biologia. Dado o seu parentesco filogenético com alguns outros flavivírus encontrados naturalmente em mosquitos de diferentes géneros colhidos em diferentes regiões do globo, o CFAV poderá ser usado como modelo genérico para o estudo de ISF, incluindo a análise da sua estratégia replicativa, e as interações que estabelece, *in vitro* e *in vivo*, com as células em que replica. Tal pode acontecer, ou pelo menos algumas das etapas deste processo só podem ser avaliadas, se estiverem disponíveis ferramentas, tais como, anticorpos, de preferência monoclonais. Obviamente, que a produção destes anticorpos pode ser levado a cabo usando péptidos sintéticos como antígenos, a verdade é que, o processo funciona melhor se forem usadas proteínas inteiras, devido a grande variedade de epitopos disponíveis. Dada a pouca informação disponível sobre a biologia dos ISF em geral, bem como do CFAV em particular, seria também, interessante analisar a estrutura tridimensional das suas proteínas, por exemplo, através de técnicas cristalográficas, ou estudar a actividade de algumas delas, tais como, as não estruturais (NS), e para tal, é necessário que a possamos trabalhar com proteínas pura ou, pelo menos, extractos proteicos com elevado grau de pureza.

Neste contexto, e tendo em conta o facto de haver enorme limitação no que respeita à disponibilidade de ferramentas básicas que possam ser usadas, laboratorialmente, tendo em vista uma melhor compreensão da sua biologia, com o presente trabalho propusemo-nos a otimizar protocolos que permitissem a expressão e purificação de quatro proteínas codificadas pelo CFAV: duas proteínas estruturais (C e E), e duas proteínas não estruturais (NS3hel e NS5B). Ainda que se as proteínas em causa sejam codificadas por um vírus que infecta células eucariontes, a verdade é que as vantagens dos sistemas de expressão proteica baseados na utilização de *E. coli* levaram

à sua escolha como hospedeira para a produção de proteínas virais como fusões a “caudas” (tags) de hexahistidina.

A primeira etapa do trabalho aqui apresentado consistiu na expansão do CFAV a partir de um lisado gentilmente cedido por V. Stollar numa linha celular, designada C6/36, de forma a obter uma suspensão viral de elevado título. Tal como descrito inicialmente por Stollar e Thomas (1975), foi possível observarmos o efeito citopático que caracteriza a replicação deste vírus nestas células, o qual se traduz pela formação de sincícios de tamanhos variáveis. Ainda que não tenhamos procedido à titulação da suspensão viral assim obtida (através da quantificação de ácidos nucleicos ou antígenos virais), o RNA extraído de sobrenadantes de culturas permitiu verificarmos a especificidade dos *primers* desenhados no decurso deste trabalho, através da amplificação de sequências virais presentes no cDNA que foi preparado a partir de extractos de RNA viral. Estes mesmos *primers* foram utilizados na tentativa de amplificação de sequências virais a partir de DNA total também extraído dos sedimentos celulares (células infectadas). Os resultados obtidos evidenciam que durante a sua replicação em células C6/36, o CFAV não se integrou no genoma das mesmas. De igual modo, os *primers* desenhados parecem ser específicos para o CFAV (como esperado) não permitindo a amplificação dos fragmentos remanescentes de um vírus (agora defetivo) designado *Cell Silence Agent* (CSA), integrado sob a forma de múltiplos fragmentos de DNA no genoma das células C6/36 (Crochu *et al.*, 2004; Hoshino *et al.*, 2007). No entanto, foi possível constatar a presença de um fragmento de DNA com tamanho de aproximadamente 200pb aquando da utilização de *primers* complementares à sequência codificante de NS5B. Estes resultados não são, no entanto inesperados, dado que os *primers* utilizados nas amplificações de parte da sequência codificante de NS5B têm como alvo uma região conservada da região que codifica esta proteína, presente no genoma de todos os flavivírus conhecidos (Parreira *et al.*, 2012).

A obtenção de cDNA a partir do RNA viral, através da síntese por transcrição reversa (RT), permitiu, então, a amplificação, por PCR, das regiões dos genes codificantes das proteínas C (proteína estrutural da cápside), E (glicoproteína estrutural de superfície), NS3hel (domínio que confere à proteína não estrutural NS3 actividade de helicase de RNA) e NS5B (polimerase de RNA dependente de RNA) do CFAV, sob a forma de fragmentos de DNA com aproximadamente 321pb, 1180pb, 1236pb e 2660pb.

A amplificação, por PCR, destas regiões alvo foram conseguidas após optimização do protocolo de amplificação. Além disso, e porque a expressão de proteínas heterólogas com domínios hidrofóbicos pode gerar uma série de problemas, as sequências complementares aos *primers* que permitiram amplificar as sequências codificantes das proteínas estruturais C e E foram identificadas de forma a eliminar dos produtos amplificados as regiões que codificam secções hidrofóbicas destas proteínas. Desta forma, antecipamos melhorar a sua expressão em *E. coli*. De modo semelhante, o desenho dos *primers* para a amplificação de NS3hel teve por base a delimitação do domínio responsável pela sua actividade de helicase de RNA na sequência de NS3 do CFAV. De facto, embora muitos estudos já reportem a caracterização do domínio com actividade de protease deste tipo de proteínas codificadas por flavivírus diversos, muito poucos são aqueles que se dedicam ao estudo das helicases.

Neste trabalho foi proposta, *a priori*, a inserção independentemente dos produtos de PCR obtidos para as quatro regiões (C, E, NS3hel e NS5B) no vector pJTE1.2/*blunt* usando *E. coli* NovaBlue como hospedeira de clonagem. Este procedimento permitiu a estabilização das sequências codificantes de proteínas do CFAV no vector plasmídico e a obtenção de quantidades, *a priori* ilimitadas, de fragmentos de DNA que codificam as proteínas em estudo, através da sua amplificação, por PCR. Posteriormente, a expansão dos potenciais clones recombinantes em meio adequado, subsequente extracção do respectivo DNA plasmídico, confirmação do perfil de digestão enzimática de cada um dos clones recombinantes, a amplificação (por PCR), a purificação, consumou-se a reacção de ligação dos insertos obtidos aos vectores de expressão utilizados (pET-28b e pET-29b).

A utilização da bactéria Gram-negativa *E. coli* como hospedeira de expressão apresenta inúmeras vantagens tais como, rápido crescimento, fácil sub-cultura e conservação, e a rápida expressão das proteínas a produzir, as quais podem ser obtidas com, elevado rendimento e a custo relativamente baixo (Yin *et al.*, 2007; Demain e Vaishnav, 2009). Apesar, dos inúmeros sistemas de expressão de proteínas heterólogas centrados na utilização de *E. coli* apresentam, frequentemente, uma notável capacidade para produzir proteínas com elevado rendimento, eles apresentam ainda assim, várias limitações. À parte do facto de que a expressão, em células procariontes, de proteínas com origem em células eucariontes (ou nos seus vírus) carecer das eventuais

modificações pós-traducionais que estas podem sofrer *in vivo*, a expressão de proteínas heterólogas pode resultar na sua produção em quantidades diminutas devido à sua toxicidade. Em alternativa, se as sequências que codificam as proteínas a expressar contém muitos códons raramente utilizados em *E. coli*, as limitações quantitativas decorrentes da baixa quantidade de tRNAs no citoplasma do hospedeiro de expressão pode, ainda, levar à incorporação de aminoácidos erróneos e/ou o truncamento do polipéptido a ser sintetizado, causando efeitos deletérios sobre os níveis de expressão e a qualidade de proteínas heterólogas produzidas (Mirzahoseini *et al.*, 2011). Porém, e como forma a minimizar este efeito, encontram-se disponíveis diversas estirpes bacterianas geneticamente modificadas, para atender às mais diversas condições de expressão de proteínas recombinantes (Baneyx, 1999).

Neste trabalho, as estirpes de expressão utilizadas foram *E. coli* BL21(DE3)pLysS e *E. coli* Rosetta(DE3)pLysS. Esta escolha baseou-se nalgumas das suas características. Por um lado, ambas possuem um gene que codifica uma polimerase de RNA do fago T7 (sob o controlo transcricional do promotor *lacUV5*) integrado no genoma de uma variante não indutível do fago lambda (DE3), o qual pode ser induzido com IPTG. Além de apresentarem deficiência nas proteases OmpT e Lon, possuem um plasmídeo designado pLysS (ou um dos seus derivados) que codifica um gene que lhes confere resistência ao cloranfenicol e um gene que codifica a lisozima do fago T7, a qual é, por sua vez, inibidora natural da respectiva polimerase de RNA e, por consequência, permite manter o nível de expressão proteica basal a um nível reduzido. Por outro lado, a estirpe Rosetta(DE3)pLysS tem como característica muito particular transportar genes que codificam tRNA para códons raramente utilizados em *E. coli* (AGG/Arg, AGA/Arg, AUA/Ile, CUA/Leu, CCC/Pro, e GGA/Gly) num derivado de pLysS designado pRARE. Basicamente, a utilização destas duas estirpes de expressão permitiram-nos verificar até que ponto as diferenças entre as suas características podem afectar a expressão de uma, ou mais, das proteínas que nos propusemos produzir.

Quando se pretende produzir uma proteína heteróloga usando *E. coli* como hospedeira de expressão, a escolha do sistema de expressão é fundamental. Um dos sistemas de expressão mais frequentemente usado, e que inclui uma panóplia de vectores plasmídicos, é comercialmente conhecido por pET (Novagen). Os vectores pET integram uma família de plasmídeos a partir dos quais a expressão proteica pode

ser induzida, e controlada, através da acção combinada da polimerase de RNA do fago T7 e dos promotores de transcrição por esta reconhecidas (Snyder e Champness, 1997; Baneyx, 1999). Além disso, este sistema permite obter níveis de expressão mais elevados, maior controlo sobre o nível de expressão basal, bem como escolher as “caudas” (*tags*) de fusão em N-terminal e C-terminal para detecção, purificação e localização da proteína a ser produzida ([http://pef.aibn.uq.edu.au/support/material/download/Novagen pET system manual 2003.pdf](http://pef.aibn.uq.edu.au/support/material/download/Novagen_pET_system_manual_2003.pdf)).

Os vectores de expressão utilizados neste trabalho foram o pET-28b e o pET-29b. Estes vectores contêm um promotor reconhecido pela polimerase de RNA do fago T7 (o qual não é reconhecido pela polimerase de RNA de *E. coli*), uma sequência do operador *lac* (*lacI*), o local de clonagem múltipla e um gene que confere resistência à canamicina (*kan*).

A inserção dos vectores de expressão (pET-28b e pET-29b) contendo os insertos desejados (C, E, NS3hel e NS5B) em hospedeiros de expressão, permitiu, assim, estudar os efeitos da indução da expressão de proteínas recombinantes e avaliar a sua toxicidade através da monitorização do crescimento bacteriano, utilizando duas temperaturas 30 e 37°C. Os resultados obtidos para os derivados de *E. coli* BL21(DE3)pLysS contendo os 4 plasmídeos de expressão construídos (induzidas a 30 e 37°C), mostraram que a indução da expressão proteica não pareceu ter qual quer impacto no crescimento das mesmas. O mesmo não aconteceu com as culturas de *E. coli* Rosetta(DE3)pLysS contendo o plasmídeo de expressão da proteína C e o plasmídeo de expressão da proteína E, em que uma vez induzida a expressão proteica e incubadas à 37°C, as culturas bacterianas apresentaram uma redução na sua taxa de crescimento, quando comparadas com as culturas bacterianas não induzidas, especialmente evidente no caso proteína viral E. Por outro lado, as mesmas culturas quando induzidas e incubadas à 30°C apresentam uma ligeira redução na taxa de crescimento quando comparadas com as culturas não induzidas. Estes resultados sugerem que a desaceleração no crescimento bacteriano (relativamente às estirpes onde a expressão proteica não foi induzida) possa resultar ou de um qualquer efeito tóxico associado à produção das proteínas heterólogas em questão ou, em alternativa, da mobilização da maquinaria celular para a produção das proteínas do CFAV, reduzindo, assim, o crescimento bacteriano. Miroux e Walker, 1996, demonstraram que um componente essencial da toxicidade da sobre-expressão de



proteína em sistemas de *E. coli* pode estar ligado a transcrição do gene de interesse e que, a sobre-expressão de certos produtos, como por exemplo, da  $\beta$ -galactosidase, pode levar a destruição de RNAs ribossomais e, subsequentemente, causar efeitos letais à célula. Outros estudos evidenciam que a sobre-expressão, provavelmente, ocorre devido ao desacoplamento da transcrição da tradução (Iost e Dreyfus, 1995; Makarova *et al.*, 1995). No entanto, uma das estratégias utilizada para suprimir os efeitos tóxicos provocados pelo nível de expressão basal é a co-transformação de um plasmídeo (pLysS ou pLysE) que codifica a lisoizima T7 (inibidora da polimerase de RNA do fago T7) com o plasmídeo contendo o gene de interesse (Miroux e Walker, 1996; Mierendorf *et al.*, 1999).

Ao contrário das culturas de *E. coli* Rosetta(DE3)pLysS contendo o plasmídeo de expressão da proteína C e o plasmídeo de expressão da proteína E, as estirpes de *E. coli* Rosetta(DE3)pLysS contendo o plasmídeo de expressão da proteína NS3hel e o plasmídeo de expressão da proteína NS5B, quando induzidas e incubadas à 30 e 37°C, parecem não exercer qualquer efeito de toxicidade com impactos directos sobre o metabolismo celular, pois que, apresentaram um comportamento similar, quando comparadas com as culturas transformadas (não induzidas) e com as culturas não transformadas (induzidas e não induzidas). Além disso, estes dados dão a indicação de que estas proteínas estão, de facto, a ser produzidas *in vivo*.

Quanto ao nível de expressão proteica, foi possível constatar, a aparente ausência, 4 horas após indução a 30 e 37°C, das proteínas desejadas nos extractos proteicos totais obtidos a partir de derivados de *E. coli* BL21(DE3)pLysS contendo os plasmídeos construídos. Por outro lado, nos extractos proteicos totais de derivados de *E. coli* Rosetta(DE3)pLysS contendo o plasmídeo de expressão da proteína C, 4 horas após indução a 37°C com 1 mM de IPTG, foi possível constatar a presença de uma banda com aproximadamente 12kDa, correspondente à massa estimada da proteína C. Mas, o mesmo nem sempre foi possível constatar nos extractos proteicos totais derivados de *E. coli* Rosetta(DE3)pLysS contendo os 3 plasmídeos de expressão das proteínas E, NS3hel e NS5B. Estes dados corroboram com um estudo desenvolvido em parceria com vários laboratórios em que confirmaram que, o uso de estirpes que transportam o plasmídeo pLysSRARE (codifica tRNA para códons raramente usados por *E. coli*) apresentou melhores resultados, aquando da expressão de proteínas de origem em

células eucariontes (Berrow *et al.*, 2006). Estes resultados permitem comparar, assim, o nível de expressão de proteínas desejadas nos dois sistemas utilizados, que diferem num facto em particular, pLysS [*E. coli* BL21(DE3)pLysS] e pLysSRARE [*E. coli* Rosetta (DE3)pLysS].

Tanto a temperatura de crescimento bacteriano, como a concentração de indutor (IPTG), podem influenciar, independentemente, o crescimento celular e a produção de proteína. Estudos publicados têm demonstrado que a temperatura de crescimento bacteriano pode influenciar na produção de proteína, não só acelerando a sua síntese através da estimulação, dentro de certos limites, do metabolismo celular, como, e principalmente, afectando a sua solubilidade. Normalmente, a temperaturas baixas a solubilidade da proteína aumenta, já a temperaturas mais elevadas a sua produção em maior escala pode conduzir à formação de agregados insolúveis, também designados de corpos de inclusão (Carrió e Villaverde, 2002; Berrow *et al.*, 2006; Vera *et al.*, 2007). Um outro estudo referiu que concentrações elevadas de indutor podem resultar em rendimentos de proteína heteróloga superior a 50% de proteína celular total e que, a incapacidade da bactéria suportar todas as modificações pós-traducionais que uma proteína requer para a aquisição de uma estrutura tridimensional correcta podem proporcionar também a formação de corpos de inclusão (Baneyx e Mujacic, 2004).

No presente estudo foram verificados os níveis de expressão proteica através da avaliação de dois parâmetros distintos: a temperatura de crescimento bacteriano e a concentração do indutor (IPTG). Das três temperaturas testadas (15, 25 e 37°C durante 16h, com 1 mM de IPTG) só foi possível verificar que houve expressão da proteína C, nos extractos proteicos totais derivados de *E. coli* Rosetta(DE3)pLysS contendo o plasmídeo de expressão da proteína C, induzidos a 25 e 37°C. Além disso, não foram visualizadas de forma clara as bandas correspondentes as proteínas E, NS3hel e NS5B nos extractos proteicos totais derivados de *E. coli* Rosetta(DE3)pLysS contendo os plasmídeos de expressão destas mesmas proteínas. Quanto as duas concentrações de IPTG testadas, 0,4 mM e 1 mM, os resultados foram equivalentes e, apenas, foi possível visualizar uma banda com aproximadamente 12kDa, correspondente à massa estimada para a proteína C, em ambas as concentrações. Muitas vezes, ao utilizarmos concentrações elevadas de indutor, o sistema pode responder negativamente resultando na obtenção de quantidades diminutas de proteína do que quando utilizamos

concentrações baixas de indutor. Neste caso, e ainda que os resultados obtidos, referentes às duas concentrações de IPTG utilizadas fossem equivalentes, 1 mM de IPTG não pareceu induzir qualquer alteração negativa na expressão proteica. Por esta razão, optou-se por trabalhar, nas fases seguintes, 1 mM de IPTG, de acordo com as instruções referidas no manual de referência [http://pef.aibn.uq.edu.au/support/material/download/Novagen\\_pET\\_system\\_manual2003.pdf](http://pef.aibn.uq.edu.au/support/material/download/Novagen_pET_system_manual2003.pdf)). Quanto, a temperatura, optou-se por trabalhar com a temperatura de 37°C e induzir as culturas durante 4 horas.

Como não existem regras pré-definidas no que diz respeito à definição de condições que permitam a expressão heteróloga de proteínas num qualquer hospedeiro, tanto nas etapas de expressão como durante a subsequente purificação, os protocolos utilizados são quase sempre adaptações dependentes da proteína a ser produzida. Além disso, a questão da solubilidade da proteína recombinante é um factor determinante na elaboração dos protocolos de purificação.

Neste trabalho, uma das limitações que constatamos prendeu-se com aparente incapacidade, nalguns casos, de promoção, por sonicação, de uma lise das células bacterianas, provavelmente associada à sobressaturação da suspensão bacteriana. Após a realização da sonicação, para a desintegração das células, verificou-se a solubilidade dos extractos proteicos totais obtidos, para cada uma das proteínas em estudo. Os resultados da eletroforese em gel de poliacrilamida (Tricina-SDS-PAGE) permitiram a visualização de bandas com aproximadamente 12, 43 e 49kDa, correspondentes às proteínas C, E e NS3hel, na fracção insolúvel, dando indicação de que nas estirpes onde a sua produção foi induzida tenha ocorrido a formação de corpos de inclusão. No que diz respeito aos extractos proteicos totais correspondente a proteína NS5B, não foi possível visualizar uma banda no gel, tanto na fracção solúvel como na fracção insolúvel. Tal facto pode acontecer, provavelmente, devido à sobreposição da banda correspondente a proteína recombinante com outra banda de *E. coli* Rosetta(DE3)pLysS, e/ou devido a sua produção em quantidades diminutas, que relegam a sua detecção para a utilização de métodos analíticos mais sensíveis. Analogamente, um estudo demonstra que proteínas recombinantes de CFAV são, normalmente, produzidas como corpos de inclusão insolúveis (Peyrefitte *et al.*, 2007). No entanto, nos extractos proteicos totais correspondentes as proteínas C e E, constatou-se que a lise celular não foi completa, daí a ausência de proteínas nas fracções solúveis.

Estes dados parecem indicar que, ou utilizou-se células “a mais”, e/ou as condições de sonicação não foram ideais. Consequentemente, devido ao factor tempo, não foram realizados mais protocolos de lise, além disso, queríamos verificar se as proteínas desejadas estavam, efectivamente, ou não, nos extractos proteicos totais obtidos. Obviamente, que em futuros trabalhos teríamos de fazer alterações aos protocolos utilizados, e/ou utilização de ciclos de congelação e descongelação antecedendo a sonicação, ou submetendo as células bacterianas a mecanismos de lise física mais agressivos que a sonicação, tais como o seu tratamento numa prensa de French.

Existem inúmeras técnicas disponíveis para a purificação de proteínas a partir de um extracto complexo. Os procedimentos em causa e a ordem dos métodos a aplicar dependem do tipo de proteína e empregam-se de acordo com as características da mesma. Ou seja, depende da localização e da solubilidade da proteína, da acessibilidade da “cauda” (*tag*) de fusão, e/ou se a actividade biológica deve ser mantida. O processo de purificação pode ser feito em condições desnaturantes ou nativas. Proteínas que permanecem solúveis no citoplasma, ou que são secretadas, geralmente, podem ser purificadas em condições nativas. A purificação de proteínas solúveis é de execução rápida e menos dispendiosa, do que a purificação e renaturação de proteínas a partir de corpos de inclusão (Sorensen e Mortensen, 2005). No entanto, a purificação em condições nativas pode ser prejudicada se a proteína alvo for insolúvel, agregada em corpos de inclusão, ou se possuir uma estrutura terciária que obstruí a cauda de fusão da proteína. Em situações como estas, a proteína pode ser purificada em condições em que se promova a sua desnaturação, utilizando para tal elevadas concentrações de reagentes caotrópicos, tais como, a ureia ou o hidrocloreto de guanidina, de forma, a permitir a desnaturação e a solubilização da mesma (Bornhorst e Falke, 2000; Singh e Panda, 2005).

Através da passagem de extractos proteicos totais através de uma coluna de cromatografia empacotada com resina contendo iões níquel imobilizados realizámos a purificação das proteínas C, E, NS3hel e NS5B do CFAV, expressas em fusão com uma cauda de hexahistidina, quer em condições nativas, quer em condições desnaturantes. Relativamente aos resultados obtidos, as tentativas de purificação de proteínas em condições desnaturantes revelaram a presença das proteínas com massas moleculares compatíveis com as calculadas (teoricamente) para C, E, NS3hel e NS5B. Estas

apresentavam um nível de purificação considerável, revelado pela presença de bandas proteicas bem evidentes com massas moleculares esperadas de aproximadamente 12, 43, 49 e 100kDa, na Figura 3.17.B. Quanto aos resultados obtidos na sequência da purificação de proteínas em condições nativas, estes revelaram, apenas, a presença de uma banda correspondente a uma proteína com uma massa compatível com a estimada para a proteína C. Estes resultados parecem indicar que, apenas, a proteína C pode ser purificada nas duas condições, e que as restantes proteínas só podem ser purificadas em condições desnaturantes. Face aos objectivos traçados, seria mais interessante purificarmos as proteínas desejadas em condições nativas pois que, assim, seria possível estudarmos actividade de algumas delas, nomeadamente, nomeadamente, da NS3hel e NS5B. Obviamente, que será necessário, em futuros trabalhos, ajustar alguns parâmetros de modo, a melhorar a solubilidade das proteínas obtidas. A formação de corpos de inclusão representa um obstáculo para a produção de proteínas, uma vez que, mesmo favorecendo o elevado rendimento de produto e a possibilidade de se obter o produto quase puro, a recuperação *in vitro* da sua bioactividade depende tecnicamente de diferentes e complexos procedimentos de renaturação (Carrió e Villaverde, 2002; Singh e Panda, 2005). Muitas vezes, o rendimento global de proteína bioactiva a partir de corpos de inclusão é cerca de 15-25% de proteína total e representa o maior custo de produção de proteína recombinante em *E. coli* (Singh e Panda, 2005).

Subsequentemente, com a utilização de métodos analíticos tais como a detecção com anticorpos anti-histidina (*Western Blot*), foi possível confirmar a autenticidade das proteínas produzidas e purificadas (condições nativas e condições desnaturantes) com “caudas” de histidinas. Os resultados dos extractos proteicos totais obtidos em condições desnaturantes, permitiram confirmar, de forma específica, a presença de domínios ricos em histidinas em quatro proteínas com aproximadamente 12, 43, 49 e 100kDa, as quais, e tal como esperado, deverão corresponder as proteínas C, E, NS3hel e NS5B. Pelo contrário, nos extractos proteicos obtidos em condições nativas, foi possível detectar, apenas, a proteínas C, com aproximadamente 12kDa, na Figura 3.18.B.

Feita a imunodeteção de domínios ricos em histidina em quatro proteínas, procedeu-se a quantificação de, apenas, a proteína C nos extractos totais obtidos. O valor para o extracto proteico total que contém a proteína C (1,25mg/ml), parece

aproximar-se à real quantidade de C no extracto purificado. É de salientar que mesmo nos extractos proteicos totais quando observamos directamente o extracto de proteínas totais a banda correspondente a C foi sempre facilmente identificada em todos os géis de SDS-PAGE preparados. Embora todos os demais extractos apresentem um nível de purificação aparentemente bom, as concentrações aparentes calculadas parecem estar francamente inflacionadas, tal como demonstram os resultados das quantificações efectuadas (não apresentados), devendo tal facto estar associado à presença de proteínas contaminantes.

Em suma, podemos afirmar que foi possível expressar as quatro proteínas codificadas pelo CFAV em *E. coli*. Rosetta(DE3)pLysS, sob a forma de corpos de inclusão. Apesar de os corpos de inclusão serem resistentes a proteólise, de conterem grandes quantidades de material relativamente puro (Baneyx e Mujacic, 2004; Yin *et al.*, 2007), e de simplificar a purificação de proteínas, não há qualquer garantia de que a sua renaturação *in vitro*, através, por exemplo, da remoção do agente desnaturante por diálise possa conduzir à obtenção de grandes quantidades de produto biologicamente activo (Terpe, 2006). No entanto, é do conhecimento que a extensão de agregação de proteínas é determinada, pelo menos parcialmente, por uma combinação de parâmetros do processo, incluindo a composição de meio de cultura, a temperatura de crescimento, a taxa de produção (como resultado de diversos factores, tais como a dosagem do gene alvo, a força do promotor, a estabilidade do mRNA e uso de codões), e a disponibilidade de moléculas chaperones (Baneyx, 1999; Stevens, 2000; Sorensen e Mortensen, 2005; Terpe, 2006; Ventura e Villaverde, 2006). Todos estes factores podem ser manipulados de modo melhorar a solubilidade, o rendimento, assim como, a qualidade da proteína heteróloga produzida.

## ***5. CONCLUSÕES E PERSPECTIVAS FUTURAS***

---

A expressão de proteínas e péptidos heterólogos em *E. coli* está bem documentada e é muitas vezes eficaz para a produção de quantidades suficientes de proteínas recombinantes. Diversos factores podem ser optimizados, como a estirpe hospedeira, a escolha do vector, o meio e a temperatura de crescimento, de modo a melhorar, de forma significativa, a qualidade conformacional, o rendimento de produção e a solubilidade de proteínas recombinantes produzidas (Stevens, 2000).

No presente trabalho quatro proteínas, designadas C, E, NS3hel e NS5B, codificadas pelo genoma do CFAV foram clonadas em vectores de expressão proteica (pET-28b e pET-29b) para permitir a sua expressão como proteínas de fusão com uma “cauda” de hexahistidina na região C-terminal, produção esta regulada por um promotor indutível, derivado do genoma do fago T7. A expressão de proteínas de interesse foi realizada em culturas de *E. coli* BL21 (DE3)pLysS e Rosetta(DE3)pLysS com 1mM de IPTG durante 4 horas a 37°C. A purificação das quatro proteínas em questão foi levada a cabo em condições desnaturantes e nativas, tendo a primeira abordagem permitido a obtenção de extractos proteicos parcialmente puros. Posteriormente, foi verificada a identidade das proteínas expressas usando a técnica de *Western Blot*, na qual foram detectadas, de forma específica, a presença de domínios ricos de histidina. As quatro proteínas com aproximadamente 12, 43, 49 e 100kDa, deverão, e tal como esperado, corresponder as proteínas C, E, NS3hel e NS5B.

Para além da análise de sequências de “flavivírus específicos de insectos”, amplificadas a partir de amostras colhidas em populações naturais de mosquitos de diversas regiões do mundo, bem como a análise filogenética a partir de fragmentos correspondente, maioritariamente, às regiões codificantes das proteínas E, NS3 e NS5 (Cook et al., 2012; Cook *et al.*, 2006; Crabtree *et al.*, 2003; Crochu *et al.*, 2004; Hoshino et al., 2007), a sua comparação com as de outros flavivírus, pouco tem sido extensivamente estudado no que diz respeito à expressão e purificação de proteínas, codificadas por estes genomas virais. A análise do CFAV, não é excepção, e, à semelhança do que diz respeito aos demais IFS, a literatura que reporta a análise deste vírus e dos seus produtos é extremamente limitada.

No entanto, devemos salientar que os resultados obtidos no presente trabalho deixam em aberto vários desafios, como o de optimizar processos de renaturação *in vitro*, das proteínas obtidas sob a forma de corpos de inclusão, podendo resultar em



proteínas nativas que, posteriormente, poderão vir a ser utilizadas tanto como antígenos para a produção de anticorpos, cristalizá-las para a determinação das suas estruturas tridimensionais e/ou, e especialmente no que diz respeito a NS3hel e NS5B, utilizá-las em ensaios que visem a sua caracterização bioquímica do ponto de vista da sua actividade enzimática.

## ***6. REFERÊNCIAS BIBLIOGRÁFICAS***

---

- Anfinsen, C.B.** 1973. Principles that govern the folding of protein chains. *Science*. **181**: 223–230.
- Assenberg, R., E. Mastrangelo, T. S. Walter, A. Verma, M. Milani, R. J. Owens, D. I. Stuart, J. M. Grimes, and E. J. Mancini.** 2009. Crystal Structure of a Novel Conformational State of the Flavivirus NS3 Protein: Implications for Polyprotein Processing and Viral Replication. *Journal of Virology*. **83**:12895–1290.
- Baneyx, F.** 1999. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*. **10**:411–421.
- Baneyx, F., and M. Mujacic.** 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*. **22**:1399–1408.
- Blitvich, B. J., M. Lin, K. S. Dorman, V. Soto, E. Hovav, B. J. Tucker, M. Staley, K. B. Platt, and L. C. Bartholomay.** 2009. Genomic Sequence and Phylogenetic Analysis of *Culex* flavivirus, an insect-specific flavivirus, Isolated from *Culex pipines* (Diptera: Culicidae) in Iowa. *Journal of Medical Entomology*. **46**:934–941.
- Berrow, N. S., K. Bussow, B. Coutard, J. Diprose, M. Ekberg, G. E. Folkers, N. Levy, V. Lieu, R. J. Owens, Y. Peleg, C. Pinaglia, S. Quevillon-Cheruel, L. Salim, C. Scheich, R. Vincentellic, and D. Bussow.** 2006. Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallographica*. **62**:1218–1226.
- Bollati, M., K. Alvarez, R. Assenberg, C. Baronti, B. Canard, S. Cook, B. Coutard, E. Decroly, X. De Lamballerie, E. A. Gould, G. Grard, J. M. Grimes, R. Hilgenfeld, A. M. Jansson, H. Malet, E. J. Mancini, E. Mastrangelo, A. Mattevi, M. Milani, G. Moureau, J. Neyts, R. J. Owens, J. Ren, B. Selisko, S. Speroni, H. Steuber, D. I. Stuart, T. Unge, and M. Bolognesi.** 2010. Structure and functionality in flavivirus NS-protein: Perspectives for drug design. *Antiviral Research*. **87**:125–148.
- Bornhorst, J. A., and J. F. Joseph.** 2000. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol*. **326**:245–254.

- Borowski, P., A. Niebuhr, O. Mueller, M. Bretner, K. Felczak, T. Kulikowski, and H. Schmitz.** 2001. Purification and characterization of West Nile Virus nucleoside triphosphatase (NTPase)/helicase: evidence for dissociation of the NTPase and helicase activities of the enzyme. *Journal of Virology*. **75**:3220–3229.
- Borowski, P., M. V. Heising, I. B. Miranda, C. Liao, J. Choe, and A. Baier.** 2008. Viral NS3 helicase activity is inhibited by peptides reproducing the Arg-rich conserved motif of the enzyme (motif VI). *Biochemical Pharmacology*. **76**:28–38.
- Bross, P. and N. Gregersen,** 2003. *Methods in Molecular Biology - Protein Misfolding and Disease: Principles and Protocols*. Vol. (232). Humana Press Inc.
- Cammisa-Parks, H., L.A. Cisar, A. Kane, and V. Stollar.** 1992. The complete nucleotide sequence of cell fusing agent (CFA): homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. *Virology* **189**:511–524.
- Carrió, M. M., and A. Villaverde.** 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*. **96**:3–12.
- Cook, S., G. Moureau, R. E. Harbach, L. Mukwaya, K. Goodger, F. Ssenfuka, E. Gould, E. C. Holmes, and X. de Lamballerie.** 2009. Isolation of a novel species of flavivirus and a new strain of *Culex* flavivirus (*Flaviviridae*) from a natural mosquito population in Uganda. *Journal of General Virology*. **90**:2669–2678.
- Cook, S., S. N. Bennett, E. C. Holmes, R. De Chesse, G. Moureau, and X. de Lamballerie.** 2006. Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. *Journal of General Virology*. **87**:735–748.
- Cook, S., G. Moureau, A. Kitchen, E. A. Gould, X. de Lamballerie, E. C. Holmes, and R.E. Harbach.** 2012. Molecular evolution of the insect-specific flaviviruses. *Journal of General Virology*. **93**:223–234.
- Crabtree, M. B., R. C. Sang, V. Stollar, L. M. Dunster, and B. R. Miller.** 2003. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. *Archives of Virology*. **148**:1095–1118.

- Crabtree, M. B., P. T. Nga, and B. R. Miller.** 2009. Isolation and characterization of a new mosquito flavivirus, Quang Binh virus, from Vietnam. *Archives of Virology*. **154**:857–860.
- Crochu, S., S. Cook, H. Attoui, R. N. Charrel, R. D. Chesse, M. Belhouchet, J. Lemasson, P. D. Micco, and X. de Lamballerie.** 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. *Journal of General Virology*. **85**:1971–1980.
- Chung, C. T., S. L. Niemela, and R. H. Miller.** 1989. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences*. **86**:2172–2175.
- de Lamballerie, X., S. Crouch, F. Billoir, J. Nesyts, P. de Micco, E. C. Holmes, and E. A. Gould.** 2002. Genomic sequence analysis of Tamana bat virus and its relationship with the genus *Flavivirus*. *Journal of General Virology*. **83**: 2443–2454.
- Demain, A. L., and P. Vaishnav.** 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*. **27**:297–306.
- Farfan-Ale, J. A., M. A. Loroño-Pino, J. E. Garcia-Rejon, E. Hovav, A. M. Powers, M. Lin, K. S. Dorman, K. B. Platt, L. C. Bartholomay, Soto, B. J. Beaty, R. S. Lanciotti, and B. J. Blitvich.** 2009. Detection of RNA from a novel West Nile-like virus and high prevalence of an insect-specific flavivirus in mosquitoes in the Yucatan Peninsula of Mexico. *American Journal of Tropical Medicine and Hygiene*. **80**:85–95.
- Farfan-Ale, J. A., M. A. Loroño-Pino, J.E. Garcia-Rejon, V. Soto, M. Lin, M. Staley, K. S. Dorman, L. C. Bartholomay, E. Hovav, and B. J. Blitvich.** 2010. Detection of flaviviruses and orthobunyaviruses in mosquitoes in the Yucatan Peninsula of Mexico in 2008. *Vector Borne Zoonotic Dis* **10**:777–783.
- Ferrari, E., J. wright-Minogue, J.W. Fang, B. M. Baroudy, J. Y. Lau, and Z. Hong.** 1999. Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase expressed in *Escherichia coli*. *Journal of Virology*. **73**:1649–1654.

- Geiss B. J., H. Stahla, A. M. Hannah, A. M. Gari, and S. M. Keenan.** 2009. Focus on flaviviruses: current and future drug targets. *Future Medicinal Chemistry*. 1:327-44.
- Gregersen, N., et al.** 2006. Protein Misfolding and Human Disease. *Annual Review of Genomics and Human Genetics*. 7:103-124.
- Gritsun, T. S., and E. A. Gould.** 2006. The 3' untranslated regions of *Kamiti River Virus* and *Cell Fusing Agent Virus* originated by self-duplication. *Journal of General Virology*. 87:2615–2619.
- Guyatt, K. J., E. G. Westaway, and A. A. Khromykh.** 2001. Expression and purification of enzymatically active recombinant RNA-dependente RNA polymerase (NS5) of the flavivirus Kunjin. *Journal of Virology Methods*. 92:37–44.
- Hannig, G., and S. C. Makrides.** 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology*. 16:54–60.
- Haider, S. R., H. J. Reid, and B. L. Sharp.** 2010. Modification of tricine-SDS-PAGE for online and offline analysis of phosphoproteins by ICP-MS. *Analytical and Bioanalytical Chemistry*. 397:655–664.
- Hartl, F.U. and M. Hayer-Hartl.** 2009. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*, 2009. 16:574-581.
- Hiniker, A. and Bardwell, J. C. A.** (2003). Disulfide Bond Isomerization in Prokaryotes. *Biochemistry*, 42:1180-1185.
- Hoshino, K., H. Isawa, Y. Tsuda, K. Yano, T. Sasaki, M. Yuda, T. Takasaki, M. Kobayashi, and K. Sawabe.** 2007. Genetic characterization of a new insect flavivirus isolated from *Culex pipiens* mosquito in Japan. *Virology* 359:405–414.
- Hoshino, K., H. Isawa, Y. Tsuda, K. Sawabe, and M. Kobayashi.** 2009. Isolation and characterization of a new insect flavivirus from *Aedes albopictus* and *Aedes flavopictus* mosquitoes in Japan. *Virology* 391:119–129.

- Huhtamo E, Moureau G, Cook S, Julkunen O, Putkuri N, Kurkela S, Uzcátegui NY, Harbach RE, Gould EA, Vapalahti O, de Lamballerie X.** 2012. Novel insect-specific flavivirus isolated from northern Europe. *Virology*. 433(2):471-478.
- Hu, Y.** 2005. Baculovirus as a highly efficient expression vector in insect and mammalian cells. *Acta Pharmacologica Sinica*. **26**:405–416.
- Iost, I. and M. Dreyfus.** 1995. The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO Journal*. **14**:3252–3261.
- Jaiswal, S., N. Khanna, and S. Swaminathan.** 2004. High-level expression and one-step purification of recombinant dengue virus type 2 envelope domain III protein in *Escherichia coli*. *Protein Expression and Purification* **33**:80–91.
- Jones, C. T., L. Ma, J. W. Burgner, T. D. Groesch, C. B. Post, and R. J. Kuhn.** 2003. Flavivirus cápside is a dimeric alpha-helical protein. *Journal of Virology*. **77**:7143–7149.
- Kapoor A., P. Simmonds, G. Gerold, N. Qaisar, K. Jain, J. A. Henriquez, C. Firth, D. L. Hirschberg, C. M. Rice, S. Shields, and W. I. Lipkin.** 2011. Characterization of a canine homolog of hepatitis C virus. *Proceedings of the National Academy of Sciences*. **108**:11608–11613.
- Kihara, Y., T. Satho, Y. Eshita, K. Sakai, A. Kotaki, T. Takasaki, Y. Rongsriyam, N. Komalamisra, R. Srisawat, P. Lapcharoen, S. Sumroiphon, S. Iwanaga, H. Ushijima, D. Endoh, T. Miyata, A. Sakata, N. Kashige, F. Miake, S. Fukushi, M. Saijo, I. Kurane, S. Morikawa, and T. Mizutani.** 2007. Rapid determination of viral RNA sequences in mosquitoes collected in the field. *Journal of Virological Methods*. **146**:372–374.
- Kim, D. Y., H. Guzman, R. Jr Bueno, J. A. Dennett, A. J. Auguste, C. V. Carrington, V. L. Popov, S. C. Weaver, D. W. Beasley, and R. B. Tesh.** 2009. Characterization of *Culex* flavivirus (*Flaviviridae*) strains isolated from mosquitoes in the United States and Trinidad. *Virology*. **386**:154–159.

- Kofler, R. M., F. X. Heinz, and C. W. Mandl.** 2002. Capsid protein C of Tick-borne Encephalitis Virus tolerates large internal deletions and is a favorable target for attenuation of virulence. *Journal of Virology*. **76**:3534–3543.
- Lindenbach, B. D., and C. M. Rice.** 1997. *Trans*-Complementation of Yellow Fever Virus NS1 Reveals a Role in Early RNA Replication. *American Society for Microbiology*. **71**:9608–9617.
- Lindenbach, B. D., and C. M. Rice.** 1999. Genetic interaction of Flavivirus nonstructural protein NS1 and NS4A as a determinant of replicase function. *Journal of Virology*. **73**:4611–4621.
- Lindenbach, B. D., H. Thiel, and C. M. Rice.** 2007. *Flaviviridae: The Viruses and Their Replication*. p.1101–1113. *Fields Virology*. Philadelphia: Lippincott-Raven.
- Mahmoud, K.** 2007. Recombinant protein production: strategic technology and a vital research tool. *Research Journal of Cell and Molecular Biology*. **1**:9–22.
- Makarova, O. V., E. M. Makarov, R. Sousa, and M. Dreyfus.** 1995. Transcribing of *Escherichia coli* genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proceedings of the National Academy of Sciences*. **92**:12250–12254.
- Makrides, S. C.** 1996. Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*. *Microbiological Reviews*. **60**:512–538.
- Mierendorf, R. C., B. B. Morris, B. Hammer, and R. E. Novy.** 1999. Expression and purification of recombinant proteins using the pET System. p.947–977. *The Nucleic Acid Protocols Handbook*. Edited by: Ralph Rapley© Humana Press Inc., Totowa, NJ.
- Miroux, B., and J. E. Walker.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*. **260**:289–298.
- Mirzahoseini, H., S. Mafakheri, N. S. Mohammadi, S. Enayati, and N. Mortazavidehkordi.** 2011. Heterologous Proteins Production in *Escherichia coli*:



- An Investigation on the Effect of Codon Usage and Expression Host Optimization. *Cell Journal (Yakhteh)*. **12**:453–458.
- Morales-Betoulle, M. E., M. L. Monzón Pineda, S. M. Sosa, N. Panella, M. R. López, C. Córdón-Rosales, N. Komar, A. Powers, and B. W. Johnson.** 2008. *Culex* flavivirus isolates from mosquitoes in Guatemala. *Journal of Medical Entomology* **45**:1187–1190.
- Murray, C. L., C. T. Jones, and C. M. Rice.** 2008. Architects of assembly: roles of Flaviviridae non-structural protein in virion morphogenesis. *Nature Reviews Microbiology*. **6**:699–708.
- Nomaguchi, M., T. Teramoto, L. Yu, L. Markoff, and R. Padmanabhan.** 2004. Requirements for West Nile Virus (-)- and (+)-strand subgenomic RNA synthesis in vitro by the viral RNA-dependent RNA polymerase expressed in *Escherichia coli*. *The Journal of Biology Chemistry*. **279**:12141–12151.
- Rai, M., and H. Padh.** 2001. Expression systems for production of heterologous proteins. *Current Science*. **80**:1121–1128.
- Roiz D., A. Vázquez, M. P. Seco, A. Tenorio, and A. Rizzoli.** 2009. Detection of novel insect flavivirus sequences integrated in *Aedes albopictus* (Diptera: Culicidae) in Northern Italy. *Journal of Virology*. **6**:93.
- Roosendaal, J., E. G. Westaway, A. Khromykh, and J. M. Mackenzie.** 2006. Regulated cleavages at the West Nile virus NS4A-2K-NS4B junctions play a major role in rearranging cytoplasmic membranes and Golgi trafficking of the NS4A protein. *Journal of Virology*. **80**:4623–4632.
- Parreira R., S. Cook, A. Lopes, A. P. de Matos, A. P. G. de Almeida, J. Piedade, and A. Esteves.** 2012. Genetic characterization of na insecto-specific flavivirus isolated from *Culex theileri* mosquitoes collected in southern Portugal. *Virus Research*. **167**(2):152–61.
- Perera, R., M. Khaliq, and R. J. Kuhn.** 2008. Closing the door on flaviviruses: entry as a target for antiviral drug design. *Antiviral Research*. **80**:11–22.

- Peyrefitte, C. N., B. A. M. Pastorino, M. Grandadam, D. Rolland, H. J. Tolou, and M. Bessaud.** 2007. Enzymatic characterization of a trypsin-like serine protease encoded by the genome of *Cell Fusing Agent Virus*. *Virus Genes*. **34**:185–191.
- Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A laboratory manual, 2<sup>nd</sup> edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sanger F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. **74**:5463–5467.
- Sampath A., and R. Padmanabhan.** 2009. Molecular targets for flavivirus drug discovery. *Antiviral Research*. **81**:6–15.
- Sampath, A., T. Xu, A. Chao, D. Luo, J. Lescar, and S. G. Vasudevan.** 2006. Structure-Based Mutational Analysis of the NS3 Helicase from Dengue Virus. *Journal of Virology*. **80**:6686–6690.
- Samsa, M. M., J. A. Mondotte, N. G. Iglesias, I. Assunção-Miranda, G. Barbosa-Lima, A. T. Da Poian, P. T. Bozza, and A. V. Gamarnik.** 2009. Dengue Virus Capsid Protein Usurps Lipid Droplets for Viral Particle Formation. *PLoS Pathogens*. **5**: e1000632.
- Sang, R. C., A. Gichogo, J. Gachoya, M. D. Dunster, V. Ofula, A. R. Hunt, M. B. Crabtree, B. R. Miller, and L. M. Dunster.** 2003. Isolation of a new flavivirus related to *Cell Fusing Agent Virus* (CFAV) from field-collected flood-water *Aedes* mosquitoes sampled from a dambo in central Kenya. *Archives of Virology*. **148**:1085–1093.
- Sangiambut, S., P. Keelapang, J. Aaskov, C. Puttikhunt, W. Kasinrerak, P. Malasit, and N. Sittisombut.** 2008. Multiple regions in dengue virus capsid protein contribute to nuclear localization during virus infection. *Journal of General Virology*. **89**:1254–1264.
- Schubert, A. M., and C. Putonti.** 2010. Evolution of the sequence composition of flaviviruses. *Infection, Genetics and Evolution*. **10**:129–136.

- Snyder, L., and W. Champness.** 2003. Molecular Genetics of Bacteria. p.179–183; 235 236; 282. ASM Press American Society for Microbiology. USA.
- Sonja, W., S. Miller, I. Romero-Brey, A. Merz, C. K. E. Bleck, P. Walther, S. D. Fuller, C. Antony, J. Krijnse-Locker, and R. Bartenschlager.** 2009. Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. *Cell Host & Microbe*. **5**:365–375.
- Sorensen, H. P., and K. K. Mortensen.** 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories*. **4**:1–8.
- Sørensen, H. P., and K. K. Mortensen.** 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*. **115**:113–128.
- Stapleton, J. T., S. Fount, A. S. Muerhoff, J. Bukh, and P. Simmonds.** 2011. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus *Pegivirus* within the family *Flaviviridae*. *Journal of General Virology*. **92**:233–246.
- Singh, S. M., and A. K. Panda.** 2005. Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*. **99**:303-310.
- Stevens, R.C.** 2000. Design of high-throughput methods of protein production for structural biology. *Structure* **8**:177–185.
- Stollar, V., and V. L. Thomas.** 1975. An agent in the *Aedes aegypti* cell line (Peleg) which causes fusion of *Aedes albopictus* cells. *Virology*. **64**:367.
- Sugrue, R. J., T. Cui, Q. Xu, J. Fu, and Y. C. Chan.** 1997. The production of recombinant dengue virus E protein using *Escherichia coli* and *Pichia pastoris*. *Journal of Virological Methods*. **69**:159–169.
- Tan, B., J. Fu, R. J. Sugrue, E. Yap, Y. Chan, and Y.H. Tan.** 1996. Recombination Dengue Type 1 Virus NS5 protein expressed in *Escherichia coli* exhibits RNA-dependent RNA polymerase activity. *Virology*. **216**:317–325.

- Terpe, K.** 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbial Biotechnology*. **72**:211–222.
- Tuplin, A., D. J. Evans, A. Buckley, I. M. Jones, E. A. Gould, and T. S. Gritsun.** 2011. Replication enhancer elements within the open reading frame of tick-borne encephalitis virus and their evolution within the Flavivirus genus. *Nucleic Acids Research*. **39**:7034–7048.
- Utama, A., H. Shimizu, S. Morikawa, F. Hasebe, K. Morita, A. Igarashi, M. Hatsu, K. Takamizawa, and T. Miyamura.** 2000. Identification and characterization of the RNA helicase activity of Japanese encephalitis virus NS3 protein. *Federation of European Biochemical Societies*. **465**:74–78.
- Vabulas, R.M., et al.** 2010. Protein Folding in the Cytoplasm and the Heat Shock Response. *Cold Spring Harbor Perspectives in Biology*. 2:12.
- Vera, A., N. González-Montalbán, A. Arís, and A. Villaverde.** 2007. The Conformational Quality of Insoluble Recombinant Proteins Is Enhanced at Low Growth Temperatures. *Biotechnology and Bioengineering*. **96**:1101–1106.
- Ventura, S., and A. Villaverde.** 2006. Protein quality in bacterial inclusion bodies. *TRENDS in Biotechnology*. **24**:179–185.
- Villordo S. M., and A.V. Gamarnik.** 2009. Genome cyclization as strategy for flavivirus RNA replication. *Virus Research*. **139**:230–239.
- Westaway, E. G., M. A. Brinton, S. Ya. Gaidamovich, M. C. Horzinek, A. Igarashi, L. Kaariainen, D. K. Lvov, J. S. Porterfield, P. K. Russell, and D. W. Trent.** 1985. Flaviviridae. *Intervirology*. **24**:183–192.
- Wickham T. J., T. Davis, R. R. Granados, M. L. Shuler, and H. A. Wood.** 1992. Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnology Progress*. **8**:391–396.

- Wu, J., A. K. Bera, R. J. Kuhn, and J. L. Simth.** 2005. Structure Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *Journal of Virology*. **79**:10268–10277.
- Yin, J., G. Li, X. Rena, and G. Herrler.** 2007. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology*. **127**:335–347.
- Youn S., H. Cho, D. H. Fremont, and MM. S. Diamond.** 2010. A short N-terminal peptide motif on flavivirus nonstructural protein NS1 modulates cellular targeting and immune recognition. *Journal of Virology*. **84**:9516–9532.
- Yu, F., F. Hasebe, S. Inoue, E. G. M. Mathenge, and K. Morita.** 2007. Identification and characterization of RNA-dependent RNA polymerase activity in recombination Japanese encephalitis virus NS5 protein. *Archives of Virology*. **152**:1859–1869.

**Referências eletrônicas:**

[http://www.ncbi.nlm.nih.gov/nuccore/NC\\_001564.1](http://www.ncbi.nlm.nih.gov/nuccore/NC_001564.1), 2010.

<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>, 2010.

[https://www.finnzymes.fi/tm\\_determination.html](https://www.finnzymes.fi/tm_determination.html), 2010.

<http://www.fermentas.com/en/support/technical-reference/phage-plasmid-dna/pjet12>, 2010.

<http://www.merck-chemicals.com/life-science-research/vector-table-novagen-pet-vector-table/cHdSb.s1O77QAAAEhPqsLd cab>, 2010.

[http://pef.aibn.uq.edu.au/support/material/download/Novagen pET system manual2003.pdf](http://pef.aibn.uq.edu.au/support/material/download/Novagen_pET_system_manual2003.pdf), 2010.